

**Post-Stop Fundamental Frequency Perturbation in  
Production and Perception of Mandarin Stop Voicing**

by

Yu-Hsiang (Roger) Lo

B. Sc., National Taiwan University, 2013

M. A., Leiden University, 2016

A THESIS SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF

**Doctor of Philosophy**

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL  
STUDIES

(Linguistics)

The University of British Columbia

(Vancouver)

June 2022

© Yu-Hsiang (Roger) Lo, 2022

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

**Post-Stop Fundamental Frequency Perturbation in Production and Perception of Mandarin Stop Voicing**

submitted by **Yu-Hsiang (Roger) Lo** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Linguistics**.

**Examining Committee:**

Kathleen Currie Hall, Associate Professor, Linguistics, UBC  
*Supervisor*

Molly Babel, Associate Professor, Linguistics, UBC  
*Supervisory Committee Member*

Márton Sós-kuthy, Associate Professor, Linguistics, UBC  
*Supervisory Committee Member*

Valter Ciocca, Professor, Audiology & Speech Sciences, UBC  
*University Examiner*

Anne-Michelle Tessier, Associate Professor, Linguistics, UBC  
*University Examiner*

Morgan Sonderegger, Associate Professor, Linguistics, McGill  
*External Examiner*

# Abstract

This dissertation examines how Mandarin-dominant Mandarin-English bilinguals use post-stop fundamental frequency (F0) in the production and perception of the stop voicing contrast in Mandarin and English. Their use of post-stop F0 is then compared with that of native English speakers. Additionally, the influence of cognitive load on the use of post-stop F0 is investigated. Along with cross-linguistic differences in the use of post-stop F0, this work foregrounds variability within participants and explores the production-perception interface on an individual level.

A corpus study and a set of parallel online production and perception experiments were devised. The results from the corpus study indicated that post-stop F0 following aspirated stops was lower than that following unaspirated stops in Mandarin. However, the data from the production experiment, in which the bilinguals read aloud words typifying the voicing contrast in stops, suggested the opposite pattern in both Mandarin and English. Furthermore, the post-stop F0 difference in English was larger as compared to Mandarin, indicating that more production weight was assigned to post-stop F0 in English than in Mandarin. The data from the perception experiment, which featured a forced-choice task, showed that the bilinguals used post-stop F0 as a cue in perceiving stops in both English and Mandarin, with higher post-stop F0 leading to more aspirated/voiceless responses, but they allocated more weight to post-stop F0 when the audio stimuli were presented as English words than as Mandarin words. When the bilinguals' post-stop F0 weights for English were compared with those of native English speakers, however, an asymmetry was revealed: even though both groups shared similar production weights, English listeners still had a higher perceptual weight than the bilinguals. With respect to cognitive load, which was induced by a concurrent visual search task, it

seemed to introduce more variability to the perceptual weights, but only for English listeners.

Overall, these results argue for a dual function of F0 in cueing phonological voicing in stops and tones across modalities in Mandarin. Furthermore, they suggest a dynamic nature of the post-stop F0 cue, which adapts to different language contexts, though this adaptability is constrained by the first language.

# Lay Summary

The goal of this dissertation is to understand how Mandarin-English bilinguals who are dominant in Mandarin and learn English as a second language use vowel-initial pitch to distinguish the difference between /p, t, k/ and /b, d, g/. Based on the data from 25 Mandarin-English bilinguals, it is found that, even though they pronounced words starting with /p, t, k/ with a higher pitch than /b, d, g/ words in both Mandarin and English, the pitch difference between /p, t, k/ and /b, d, g/ words was larger in English than in Mandarin. The same bilinguals also identified words with a higher pitch as more likely to start with /p/ than /b/, but they do so more often in English than in Mandarin. In addition, the bilinguals' performance was different from that of native English speakers. These results inform us both the flexibility and limitations in how bilinguals process speech.

# Preface

This dissertation is original work by the author, Yu-Hsiang (Roger) Lo. I wrote all chapters and computer code used in this project.

All projects and associated methods were approved by the Behavioural Research Ethics Board of the University of British Columbia (H19-03877). All data collection took place online, with resulting files stored on a server located in the University of British Columbia.

# Table of Contents

<b>Abstract</b> . . . . .	<b>iii</b>
<b>Lay Summary</b> . . . . .	<b>v</b>
<b>Preface</b> . . . . .	<b>vi</b>
<b>Table of Contents</b> . . . . .	<b>vii</b>
<b>List of Tables</b> . . . . .	<b>xiii</b>
<b>List of Figures</b> . . . . .	<b>xxii</b>
<b>Acknowledgments</b> . . . . .	<b>.xxxiii</b>
<b>Dedication</b> . . . . .	<b>xxxv</b>
<b>1 Introduction</b> . . . . .	<b>1</b>
1.1 Background . . . . .	1
1.2 The Current Study . . . . .	4
1.3 Organization of Chapters . . . . .	8
<b>2 A Corpus Study on Vowel-Onset F0 in Mandarin</b> . . . . .	<b>11</b>
2.1 Introduction . . . . .	11
2.1.1 Previous Studies on Vowel-Onset F0 in Mandarin . . . . .	13
2.1.2 Mechanisms behind Post-Stop F0 Perturbation . . . . .	19
2.1.3 F0 Normalization Strategies . . . . .	22

2.1.4	The Current Study . . . . .	23
2.2	Data and Methods . . . . .	26
2.2.1	Corpus Data . . . . .	26
2.2.2	Dataset Construction . . . . .	26
2.2.3	Statistical Analyses . . . . .	30
2.3	Results: Scaled F0 . . . . .	39
2.3.1	Scaled F0 at the Population Level . . . . .	40
2.3.2	Scaled F0 at the Individual Level . . . . .	40
2.4	Results: Standardized F0 . . . . .	42
2.4.1	Standardized F0 at the Population Level . . . . .	42
2.4.2	Standardized F0 at the Individual Level . . . . .	44
2.5	Discussion . . . . .	45
2.5.1	Summary of Results . . . . .	45
2.5.2	Reconciling the Differences . . . . .	47
2.6	Conclusion . . . . .	52
<b>3</b>	<b>The Dual Role of Post-Stop F0 in the Production and Perception of Stops in L1 Mandarin-L2 English Bilinguals . . . . .</b>	<b>54</b>
3.1	Introduction . . . . .	54
3.2	Background . . . . .	55
3.2.1	Fundamental Frequency as a Cue to Lexical Tone . . . . .	55
3.2.2	Fundamental Frequency as a Cue to Stop Voicing . . . . .	56
3.2.3	Post-Stop F0 at L1 Production-Perception Interface . . . . .	59
3.2.4	Post-Stop F0 at L2 Production-Perception Interface . . . . .	60
3.2.5	L1 Influence on L2 Cue Use . . . . .	61
3.2.6	Goals of the Current Study . . . . .	63
3.3	Production Experiment . . . . .	64
3.3.1	Participants . . . . .	64
3.3.2	Stimuli . . . . .	66
3.3.3	Procedure . . . . .	68
3.3.4	Acoustic Measurements . . . . .	69
3.3.5	Participant Inclusion Criteria . . . . .	72
3.3.6	Omitted Data . . . . .	72

3.3.7	Statistical Analyses . . . . .	72
3.3.8	Results: Production of Post-Stop F0 . . . . .	78
3.3.9	Results: Production Weights of Post-Stop F0 . . . . .	82
3.3.10	Interim Discussion: Production . . . . .	86
3.4	Perception Experiment . . . . .	86
3.4.1	Participants . . . . .	88
3.4.2	Stimuli . . . . .	88
3.4.3	Procedure . . . . .	94
3.4.4	Additional Participant Inclusion Criteria . . . . .	97
3.4.5	Omitted Data . . . . .	98
3.4.6	Statistical Analyses . . . . .	98
3.4.7	Results: Perceptual Weights of Post-Stop F0 . . . . .	101
3.4.8	Comparing Individual Post-Stop F0 Weights across Production and Perception . . . . .	106
3.5	Discussion . . . . .	109
3.5.1	Summary of Results . . . . .	109
3.5.2	Flexibility of Cue-Weighting in L2 . . . . .	112
3.5.3	Role of Tone in Post-Stop F0 . . . . .	112
3.5.4	A Trade-Off between Post-Stop F0 and Tone? . . . . .	113
3.5.5	Production-Perception Interface . . . . .	114
3.6	Conclusion . . . . .	115
<b>4</b>	<b>Differences in Post-Stop F0 in the Production and Perception of the English Stop Voicing Contrasts by L1 English and L1 Mandarin Speakers . . . . .</b>	<b>117</b>
4.1	Introduction . . . . .	117
4.1.1	Speech Learning Model (SLM) . . . . .	118
4.1.2	Perceptual Assimilation Model (PAM) . . . . .	119
4.1.3	Perceptual Interference between Consonants and Tone . . . . .	121
4.1.4	Goals of the Current Study . . . . .	123
4.2	Production Experiment . . . . .	127
4.2.1	Participants . . . . .	127
4.2.2	Stimuli, Procedure, Annotations, and Measurements . . . . .	128

4.2.3	Omitted Data . . . . .	128
4.2.4	Statistical Analyses . . . . .	128
4.2.5	Results . . . . .	130
4.2.6	Interim Discussion: Production . . . . .	137
4.3	Perception Experiment . . . . .	138
4.3.1	Participants . . . . .	138
4.3.2	Stimuli and Procedure . . . . .	138
4.3.3	Statistical Analyses . . . . .	138
4.3.4	Results . . . . .	139
4.3.5	Further Analysis: Comparing Post-Stop F0 in Production and Perception . . . . .	146
4.4	Discussion . . . . .	146
4.4.1	Summary of Results . . . . .	146
4.4.2	Individual Variation in Native and Non-Native Speech . . . . .	148
4.4.3	The Influence of L1 Perceptual Interference on L2 Stop Perception . . . . .	150
4.4.4	Production-Perception Interface . . . . .	151
4.5	Conclusion . . . . .	152
<b>5</b>	<b>The Effect of Cognitive Load on the Use of VOT and Post-Stop F0 in the Perception of Stop Voicing Contrasts by L1 English and L1 Mandarin Listeners . . . . .</b>	<b>154</b>
5.1	Introduction . . . . .	154
5.1.1	Attentional Modulation of Cue Weight . . . . .	155
5.1.2	Goals of the Current Study . . . . .	158
5.2	Perception Experiment with Cognitive Load . . . . .	161
5.2.1	Participants . . . . .	161
5.2.2	Stimuli . . . . .	162
5.2.3	Procedure . . . . .	162
5.2.4	Statistical Analyses . . . . .	166
5.2.5	Results: Visual Search Task . . . . .	167
5.2.6	Results: Perceptual Weights . . . . .	168
5.3	Discussion . . . . .	181

5.3.1	Summary of Results . . . . .	181
5.3.2	Cognitive Load Experiments: Same, Same but Different . . . . .	183
5.3.3	Language-Specific Effects of Cognitive Load? . . . . .	184
5.4	Conclusion . . . . .	186
<b>6</b>	<b>General Discussion . . . . .</b>	<b>187</b>
6.1	Summary of Main Findings . . . . .	187
6.2	Revisiting the Research Questions . . . . .	189
6.3	Reconciling the Differences between the Corpus and Experimental Studies . . . . .	191
6.4	Asymmetry between Production and Perception . . . . .	193
6.4.1	The Production-Perception Interface and L1 Influence . . . . .	193
6.4.2	Individual Differences in Production and Perception . . . . .	196
6.5	Conclusion . . . . .	202
	<b>Bibliography . . . . .</b>	<b>204</b>
<b>A</b>	<b>Language Background Questionnaire . . . . .</b>	<b>226</b>
A.1	Language . . . . .	226
A.2	You beyond Language . . . . .	229
A.3	Your Caretakers . . . . .	230
<b>B</b>	<b>Bilingual Language Profile Survey . . . . .</b>	<b>231</b>
B.1	Biographical Information . . . . .	231
B.2	Language History . . . . .	232
B.3	Language Use . . . . .	233
B.4	Language Proficiency . . . . .	234
B.5	Language Attitudes . . . . .	235
<b>C</b>	<b>Supplementary Materials for Chapter 2 . . . . .</b>	<b>236</b>
C.1	Speaker Demographic Information . . . . .	236
C.2	Summary Statistics of the Dataset . . . . .	237
C.3	Output of Models with Weighted Effect Coding . . . . .	240
C.4	Posterior Summaries of Individual-Level Parameters . . . . .	241

<b>D</b>	<b>Supplementary Materials for Chapter 3</b>	<b>250</b>
D.1	Participant Demographic Information	250
D.2	Statistical Model Specification	251
D.2.1	Production Model for Post-Stop F0	252
D.2.2	Perceptual Model for Post-Stop F0 Weight	254
D.3	Individual-Level Posterior Parameter Summaries	255
<b>E</b>	<b>Supporting Materials for Chapter 4</b>	<b>271</b>
E.1	Participant Demographic Information	271
E.2	Statistical Model Specification	272
E.2.1	Production Model for Post-Stop F0	272
E.2.2	Perceptual Model for Post-Stop F0 Weight	275
E.3	Individual-Level Posterior Parameter Summaries	277
<b>F</b>	<b>Supporting Materials for Chapter 5</b>	<b>287</b>
F.1	Participant Demographic Information	287
F.2	Accuracy of the Visual Search Task	290
F.3	Individual Guessing Probabilities	291
F.4	Posterior Summaries of Individual-Level Parameters	296
<b>G</b>	<b>Supporting Materials for Chapter 6</b>	<b>308</b>
G.1	Statistical Model Specification	308
G.2	Statistical Model Output	312

# List of Tables

Table 2.1	Summary of previous experimental studies on vowel-onset F0 perturbation in Mandarin. . . . .	16
Table 2.2	Candidate models on scaled vowel-onset F0 considered in the model comparison, with their ELPD-LOO means and standard deviations. An intercept was included in each model but is omitted here to save space. . . . .	37
Table 2.3	Candidate models on standardized vowel-onset F0 considered in the model comparison, with their ELPD-LOO means and standard deviations. An intercept was included in each model but is omitted here to save space. . . . .	38
Table 2.4	Model comparison results for key scaled-F0 model pairs, expressed in differences in ELPD-LOO and associated standard errors. Pairs that are judged to differ in predictive power are marked by asterisks. . . . .	39
Table 2.5	Marginal posterior summaries for population-level parameters from M6 (scaled F0). The contrast coding scheme for each variable is described in Section 2.2.3. The parameters whose effects are judged to be strong are marked with **, and those whose effects are judged to be weak are marked with *. . . . .	41
Table 2.6	Model comparison results for key standardized-F0 model pairs, expressed in differences in ELPD-LOO and associated standard errors. Pairs that are judged to differ in predictive power are marked by asterisks. . . . .	43

Table 2.7	Marginal posterior summaries for population-level parameters from M3 (standardized F0). The contrast coding scheme for each variable is described in Section 2.2.3. The parameters whose effects are judged to be strong are marked with **, and those whose effects are judged to be weak are marked with *.	44
Table 2.8	Comparison of outputs from models on the scaled-F0 and standardized-F0 datasets. The symbol $\approx$ indicates that there is little evidence for the vowel-onset F0 to differ between the levels on the two sides of the symbol. The symbols $<$ and $>$ mean that there is strong evidence that vowel-onset F0 is different across the levels in the specified direction. The symbol $\lesssim$ stands for weak evidence that the vowel-onset F0 associated with the level to the left of the symbol is smaller than that associated with the level on the right.	46
Table 3.1	Main findings from previous experimental studies on the post-stop F0 perturbation effect in Mandarin.	58
Table 3.2	Predicted production and perception results under difference hypotheses.	65
Table 3.3	Stimuli used in the Mandarin production experiment. The — symbol represents a consonant-vowel combination that violates the Mandarin phonotactics, and the = symbol stands for an accidental gap.	67
Table 3.4	Stimuli used in the English production experiment.	68
Table 3.5	Candidate post-stop F0 models considered in model comparison, with their ELPD-LOO means and standard deviations. An intercept was included in each model but is omitted here to save space.	76
Table 3.6	Means and standard deviations for VOT and post-stop F0 in hertz (split by gender) for the 25 L1 Mandarin-L2 English bilinguals' productions of Mandarin and English word-initial stops and sonorants.	79

Table 3.7	Model comparison results for key model pairs, expressed in differences in ELPD-LOO and associated standard errors. Pairs that are judged to differ in predictive power are marked by asterisks. . . . .	79
Table 3.8	Marginal posterior summaries for key population-level parameters from M6. The contrast coding scheme for each variable is described in Section 3.3.7. The parameters whose effects are judged to be strong are marked with **, and those whose effects are judged to be weak are marked with *. . . . .	82
Table 3.9	Candidate perceptual models considered in model comparison, with their ELPD-LOO means and standard deviations. . . . .	101
Table 3.10	Model comparison results for key perception model pairs, expressed in differences in ELPD-LOO and associated standard errors. Pairs judged to differ in predictive power are marked by asterisks. . . . .	103
Table 3.11	Marginal posterior summary for key population-level parameters from M4. The parameters whose effects are judged to be strong are marked with **, and those whose effects are judged to be weak are marked with *. . . . .	104
Table 3.12	Predicted production and perception results under difference hypotheses. . . . .	111
Table 4.1	Predicted production and perception results under different frameworks. . . . .	124
Table 4.2	Candidate vowel-onset F0 models considered in the model comparison, with their ELPD-LOO means and standard deviations. An intercept was included in each model but is omitted here. . . . .	130
Table 4.3	Means and standard deviations for VOT and vowel-onset F0 in hertz (split by gender) from L1 English and L1 Mandarin speakers' productions of English word-initial stops and sonorants. . . . .	131

Table 4.4	Model comparison results for key model pairs, expressed in differences in ELPD-LOO and associated standard errors. Pairs that are judged to differ in predictive power are marked by asterisks. . . . .	131
Table 4.5	Marginal posterior summaries for key parameters from M2. The contrast coding scheme for each variable is explained in Section 4.2.4. The parameters whose effects are judged to be strong are marked with **, and those whose effects are judged to be weak are marked with *. . . . .	134
Table 4.6	Candidate perceptual models considered in model comparison, with their ELPD-LOO means and standard deviations. . . . .	141
Table 4.7	Model comparison results for key perception model pairs, expressed in differences in ELPD-LOO and associated standard errors. Pairs judged to differ in predictive power are marked by asterisks. . . . .	142
Table 4.8	Marginal posterior summary for key parameters from M4. The parameters whose effects are judged to be strong are marked with **, and those whose effects are judged to be weak are marked with *. . . . .	143
Table 4.9	Predicted and actual results under different frameworks. . . . .	148
Table 5.1	Expected results under different hypotheses. . . . .	160
Table 5.2	Candidate models considered in the model comparison for the English between-subject CL experiment, with their ELPD-LOO means and standard errors. . . . .	169
Table 5.3	Model comparison results for the English within-subject CL experiment. The numbers show differences in ELPD-LOO and associated standard errors for key model pairs. Pairs judged to differ in predictive power are marked by asterisks. . . . .	170

Table 5.4	Marginal posterior summary for key parameters from the perceptual model M4 for the English within-subject CL experiment. The symbol $\mu$ denotes the (distribution of) population mean, and the symbol $\sigma$ denotes the (distribution of) population standard deviation across participants. . . . .	176
Table 5.5	Marginal posterior summary for key parameters from the perceptual model for the English between-subject CL experiment. The symbol $\mu$ denotes the (distribution of) population mean, and the symbol $\sigma$ denotes the (distribution of) population standard deviation across participants. . . . .	177
Table 5.6	Marginal posterior summary for key parameters from the perceptual model for the Mandarin between-subject CL experiment. The symbol $\mu$ denotes the (distribution of) population mean, and the symbol $\sigma$ denotes the (distribution of) population standard deviation across participants. . . . .	180
Table 5.7	Predicted and actual results under different hypotheses. . . . .	183
Table C.1	Speaker information. . . . .	236
Table C.2	Summary statistics for consonant duration (ms) and F0 (scaled and standardized) by utterance position, lexical tone, vowel height, place of articulation (PoA), and voicing: mean, standard deviation, and number of tokens. . . . .	237
Table C.3	Marginal posterior summaries for population-level parameters from a scaled-F0 model with weighted effect coding. The model was fit with the following structure: F0 <i>sim</i> position + height + tone + voicing + PoA + (1 + position + height + tone + voicing + PoA    speaker) + (1 + position    word). The variables <b>position</b> (UTTERANCE-INITIAL = 1, UTTERANCE-MEDIAL = $-1.10$ ), <b>height</b> (HIGH = 1, NON-HIGH = $-1.10$ ), <b>tone</b> (TONE 1 = 1, TONE 4 = $-0.39$ ), and <b>PoA</b> (LABIAL = $[1,0]$ , ALVEOLAR = $[0,1]$ , VELAR = $[-1.25, -2.80]$ ) were weighted effect coded. . . . .	240

Table C.4	Marginal posterior summaries for population-level parameters from a standardized-F0 model with weighted effect coding. The model was fit with the following structure: F0 <i>sim</i> position + height + tone + voicing + PoA + (1 + position + height + tone + voicing + PoA    speaker) + (1 + position    word). The variables <b>position</b> (UTTERANCE-INITIAL = 1, UTTERANCE-MEDIAL = -.10), <b>height</b> (HIGH = 1, NON-HIGH = -1.10), <b>tone</b> (TONE 1 = 1, TONE 4 = -.39), and <b>PoA</b> (LABIAL = [1,0], ALVEOLAR = [0,1], VELAR = [-1.25, -2.80]) were weighted effect coded. . . . .	241
Table C.5	Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level. ① = intercept, ② = utt-init - (utt-init + utt-med)/2, ③ = high - (high + non-high)/2, ④ = tone 1 - (tone 1 + tone 4)/2, ⑤ = asp - unasp, ⑥ = unasp - son. . . . .	242
Table C.6	Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level. ⑦ = labial - (labial + alveolar + velar)/3, ⑧ = alveolar - (labial + alveolar + velar)/3, ⑨ = [asp - unasp] × [high - (high + non-high)/2], ⑩ = [unasp - son] × [high - (high + non-high)/2]. . . . .	244
Table C.7	Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level. ① = intercept, ② = utt-init - (utt-init + utt-med)/2, ③ = high - (high + non-high)/2, ④ = tone 1 - (tone 1 + tone 4)/2, ⑤ = asp - unasp, ⑥ = unasp - son, ⑦ = labial - (labial + alveolar + velar)/3, ⑧ = alveolar - (labial + alveolar + velar)/3. . . . .	247

Table D.1	Demographic information of the L1 Mandarin-L2 English participants. BLP = Bilingual Language Profile Score (see Section 6.4.2); Am. = American English; Br. = British English; Ca. = Canadian English. . . . .	250
Table D.2	Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level. ① = intercept, ② = high - (high + low)/2, ③ = Eng - (tone1 + tone4)/2, ④ = tone1 - tone4, ⑤ = asp - unasp, ⑥ = unasp - son, ⑦ = [asp - unasp] × [high - (high + low)/2], ⑧ = [unasp - son] × [high - (high + low)/2]. . . .	256
Table D.3	Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level. ⑨ = [asp - unasp] × [Eng - (tone1 + tone4)/2], ⑩ = [asp - unasp] × [tone1 - tone4], ⑪ = [unasp - son] × [Eng - (tone1 + tone4)/2], ⑫ = [unasp - son] × [tone1 - tone4].	259
Table D.4	Summary of posterior distributions, in terms of mean (sd) [89% CrI], of production VOT and post-stop F0 weights in Mandarin and English for individual speakers. . . . .	263
Table D.5	Summary of posterior distributions for guessing probabilities, in terms of mean (sd) [89% CrI]. . . . .	265
Table D.6	Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M4 at the individual level.	266
Table D.7	Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M4 at the individual level.	269
Table E.1	Demographic information of the L1 English participants. Am. = American English; Br. = British English; Ca. = Canadian English. . . . .	271
Table E.2	Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M2 for L1 English speakers at the individual level. . . . .	277

Table E.3	Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M2 for L1 Mandarin speakers at the individual level. . . . .	279
Table E.4	Summary of posterior distributions for individual speakers' production post-stop F0 weights, in terms of mean (sd) [89% CrI].	281
Table E.5	Summary of posterior distributions for individual guessing probabilities, in terms of mean (sd) [89% CrI]. . . . .	281
Table E.6	Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 English listeners. . . . .	282
Table E.7	Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 Mandarin listeners. . . . .	284
Table F.1	Demographic information of the L1 English participants in the CL condition. N = non-binary; Am. = American English; Br. = British English; Ca. = Canadian English. . . . .	287
Table F.2	Demographic information of the L1 Mandarin-L2 English participants in the CL condition. BLP = Bilingual Language Profile Score (see Section 6.4.2); Am. = American English; Br. = British English; Ca. = Canadian English. . . . .	288
Table F.3	Summary of posterior distributions for guessing probabilities, in terms of mean (sd) [89% CrI]. . . . .	292
Table F.4	Summary of posterior distributions for guessing probabilities, in terms of mean (sd) [89% CrI]. . . . .	294
Table F.5	Summary of posterior distributions for guessing probabilities, in terms of mean (sd) [89% CrI]. . . . .	296
Table F.6	Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 English listeners. . . . .	297
Table F.7	Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 English listeners in the non-CL condition. . . . .	300

Table F.8	Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 English listeners in the CL condition. . . . .	302
Table F.9	Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 Mandarin listeners in the non-CL condition. . . .	304
Table F.10	Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 Mandarin listeners in the CL condition. . . . .	306
Table G.1	Marginal posterior summaries for population-level parameters from the combined perception model. . . . .	312

# List of Figures

Figure 1.1	An overview of the experiments conducted in this work, together with the composition of participants in each experiment. The same target audio stimuli are used in all perception experiments. . . . .	8
Figure 2.1	Distributions of F0-difference standard deviation, separated for female and male speakers. The vertical dashed lines mark the 95% quantile for each gender group. . . . .	29
Figure 2.2	The number of tokens each speaker contributed to the dataset, broken down by phonological voicing. . . . .	31
Figure 2.3	Scaled vowel-onset F0 values as a function of utterance position, lexical tone, vowel height, place of articulation, and phonological voicing. The number below each “box” indicates the number of tokens in each distribution. . . . .	32
Figure 2.4	Standardized vowel-onset F0 values as a function of utterance position, lexical tone, vowel height, place of articulation, and phonological voicing. The number below each “box” indicates the number of tokens in each distribution. . . . .	33
Figure 2.5	Population-level parameters from M6 (scaled F0). Inner error bars represent 89% CrIs, and outer error bars represent 95% CrIs. The posterior mean for each parameter is indicated by a dot. Note that the intercept from M6 is omitted here as it takes relatively large values, which compress the $x$ -axis to make the distributions of the other parameters hard to inspect. . . . .	41

Figure 2.6	Individual-level parameters involving the <b>voicing</b> predictor from M6 (scaled F0). Inner error bars represent 89% CrIs, and outer error bars represent 95% CrIs. The posterior mean for each parameter is indicated by a dot. . . . .	43
Figure 2.7	Population-level parameters from M3 (standardized F0). Inner error bars represent 89% CrIs, and outer error bars represent 95% CrIs. The posterior mean for each parameter is indicated by a dot. . . . .	45
Figure 2.8	Individual-level parameters involving the <b>voicing</b> predictor from M3 (standardized F0). Inner error bars represent 89% CrIs, and outer error bars represent 95% CrIs. The posterior mean for each parameter is indicated by a dot. . . . .	46
Figure 2.9	Illustration of a case where F0 standardization leads to UNASPIRATED > ASPIRATED > SONORANT, but F0 scaling leads to ASPIRATED (1.09) > UNASPIRATED (1.07) > SONORANT (1.04). . . . .	51
Figure 2.10	Comparison of normalized F0 values in terms of within-speaker ranks across the scaling and standardization strategies. For each speaker, 10 aspirated and 10 unaspirated tokens were randomly sampled and shown here. The two ranks belonging to the same tokens are connected by a line segment. . . . .	52
Figure 3.1	Examples of annotated English tokens, following the guidelines given in Section 3.3.3. . . . .	70
Figure 3.2	Examples of annotated Mandarin tokens, following the guidelines given in Section 3.3.3. . . . .	71
Figure 3.3	Standardized post-stop F0 values, normed by speaker, as a function of place of articulation, language/tone, vowel height, and phonological voicing. The number under each box represents the number of tokens in the distribution. . . . .	80

Figure 3.4	Marginal posterior summaries for key parameters involving voicing for each individual speaker. The [asp – unasp] panel shows the difference in F0 between aspirated and unaspirated stops. The [unasp – son] panel shows the difference in F0 between unaspirated stops and sonorants. The [(asp – unasp) * Eng] panel shows the further difference in F0 between aspirated and unaspirated stops in English, in comparison to Mandarin. The [(asp – unasp) * Man T1] panel shows the further difference in F0 between aspirated and unaspirated stops in Mandarin Tone 1 tokens, when compared to Tone 4 tokens. The dots denote the posterior means. The inner error bars represent 89% CrIs, and the outer error bars represent 95% CrIs. .	83
Figure 3.5	Production values for VOT and post-stop F0 in the Mandarin and English stop contrasts. Raw VOT values are plotted along the $x$ -axis, while standardized post-stop F0 values by speaker are plotted along the $y$ -axis. <b>A.</b> Raw VOT values on a linear $x$ -axis. <b>B.</b> Raw positive VOT values on a log $x$ -axis to better show the category structure. . . . .	84
Figure 3.6	Group-level production weights for VOT and post-stop F0 in each language. The embedded plot shows the same data but with the same scale along both axis, and the dashed line is $y = x$ , which represents equal production weights for both dimensions. The shaded area indicates the part enlarged in the main plot. . . . .	85

Figure 3.7	<p><b>A.</b> Individual speakers' production weights for VOT and post-stop F0. The posterior means are represented by the dots. The 89% CrIs are represented by the inner error bars, and the 95% CrIs are represented by the outer error bars. <b>B.</b> Post-stop F0 weights against VOT weights, separately for each language. The lines in the background are 100 lines of linear regression (i.e., post-stop F0 weight <math>\sim</math> VOT weight) fitted using 100 random posterior draws, separated for each language. <b>C.</b> Production VOT weights across languages. <b>D.</b> Production post-stop F0 weights across languages. In <b>C</b> and <b>D</b>, the solid lines represent 100 regression lines fit with 100 posterior draws, to show direction and uncertainty in the correlation. The dashed line is <math>y = x</math>, where the VOT weight for Mandarin equals that for English. . . . .</p>	87
Figure 3.8	<p>Schematic diagram of stimulus-creating steps for target syllables. The raw recording for <i>bil</i> was first passed as Praat Manipulation objects to adjust F0 contours and vowel duration so that the resynthesized audios resembled Mandarin <i>bil</i> and <i>bi4</i> in the citation form. The intensity of the resynthesized audios were then normalized to 75 dB before being further manipulated to create the full stimulus set. . . . .</p>	89
Figure 3.9	<p>Manipulation of target stimuli for all perception experiments. <b>A.</b> Each dot represents one stimulus, with its <math>x</math> coordinate corresponding to the voice onset time (VOT) of the initial labial stop, and its <math>y</math> coordinate to the initial F0 of the following vowel. <b>B.</b> Illustration of F0 trajectory manipulation for target syllables. Note the vowel duration in actual stimuli is not necessarily 350 ms due to the trade-off between VOT and vowel duration, which was also manipulated. The invariant parts across different tokens are shifted vertically in the figure for visual clarity only. . . . .</p>	92

Figure 3.10	Sample target stimuli. <b>A.</b> Tone 1, with VOT of 0 ms and F0 of 195 Hz. <b>B.</b> Tone 1, with VOT of 80 ms and F0 of 195 Hz. <b>C.</b> Tone 4, with VOT of 0 ms and F0 of 105 Hz. <b>D.</b> Tone 4, with VOT of 80 ms and F0 of 105 Hz. The red line in each figure represents the F0 trajectory. . . . .	93
Figure 3.11	One of possible response configurations for the Mandarin perception experiment. <b>A.</b> Target trial. <b>B.</b> Filler trial. . . . .	95
Figure 3.12	Two possible response configurations for the English perception experiment. <b>A.</b> With the responses corresponding to target audio stimuli on top. <b>B.</b> With the responses corresponding to target audio stimuli on left. . . . .	96
Figure 3.13	Illustration of trial procedure in the Mandarin perception experiment. The procedure follows a typical static starting procedure in mouse-tracking experiments, where an audio stimulus is played upon the participant's clicking on the center dot, and they have to click on an option within three seconds counting from the onset of audio stimulus presentation. . . . .	97
Figure 3.14	Line charts of Mandarin-English bilinguals' aggregated categorization of word-initial stops in each language, shown as a function of VOT, post-stop F0, and tone. <b>A.</b> With VOT on the <i>x</i> -axis. <b>B.</b> With post-stop F0 on the <i>x</i> -axis to highlight its effect on categorization. . . . .	102
Figure 3.15	Individual participants' guessing probabilities in each language. The dots represent the posterior means while the error bars stand for the 89% CrI. . . . .	105
Figure 3.16	Individuals' estimated weights from the perceptual model. <b>A.</b> Distributions of individual weights along various dimensions for Mandarin and English. <b>B.</b> Differences in cue weights along the same dimension across languages. In both figures, posterior means are represented by the dots. The 89% CrIs are marked by the inner error bars, while the 95% CrIs are marked by the outer error bars. . . . .	107

Figure 3.17	Scatter plots showing relationships (or lack thereof) between various cues. <b>A.</b> Category boundaries across languages. <b>B.</b> VOT weights across languages. <b>C.</b> Post-stop F0 weights across languages. <b>D.</b> Tone weights across languages. <b>E.</b> F0 vs. VOT in Mandarin. <b>F.</b> F0 vs. VOT in English. <b>G.</b> F0 vs. tone in Mandarin. <b>H.</b> F0 vs. tone in English. <b>I.</b> Differences in post-stop F0 weights vs. differences in tone weights. Solid lines represent 100 regression lines fit with 100 posterior draws, to show direction and uncertainty in the correlation. The dashed line in <b>A-D</b> is $y = x$ , where the intercept or VOT / post-stop F0 / tone weight for Mandarin equals that for English. . . . .	108
Figure 3.18	Post-stop F0 weightings across perception and production for each participant in Mandarin and English. Solid lines represent 100 regression lines fit with 100 posterior draws, to show direction and uncertainty in the correlation. . . . .	109
Figure 4.1	Illustration of equivalence classification between an L1 sound (i.e., /p/ in Mandarin) and an L2 sound (i.e., /b/ in English) that are phonetically similar but still auditorily different. <b>A.</b> The SLM predicts an eventual merged category that takes on both L1 and L2 phonetic properties. <b>B.</b> The PAM-L2 allows for a single phonological category with different language-specific phonetic realizations. . . . .	121
Figure 4.2	Boxplot of standardized post-stop F0 values, normed by speaker, as a function of place of articulation, language group, vowel height, and phonological voicing. . . . .	132
Figure 4.3	Population-level parameters from M2. Inner error bars represent 89% CrIs, and outer error bars represent 95% CrIs. The posterior mean for each parameter is indicated by a dot. . . . .	135
Figure 4.4	Marginal posterior summaries of parameters from M2 for participants in each language group. The dots represent the posterior means while the error bars enclose parameter values within 89% CrI. . . . .	136

Figure 4.5	Production post-stop F0 weights, as approximated by Cohen’s <i>d</i> , for L1 English and L1 Mandarin speakers. The posterior means of population-level weights are marked by the dashed line, and the shaded areas represent the 89% CrI of the mean weights. The posterior means of weights at the individual level are denoted by the dots, with individual error bars defining the 89% CrI. . . . .	137
Figure 4.6	L1 English and L1 Mandarin listeners’ aggregated categorization results of word-initial stops in English, shown as a function of VOT, post-stop F0, and tone. <b>A.</b> With VOT on the <i>x</i> -axis. <b>B.</b> With post-stop F0 on the <i>x</i> -axis to highlight its effect on categorization. . . . .	140
Figure 4.7	Population-level parameters from M4. Inner error bars represent 89% CrIs, and outer error bars represent 95% CrIs. The posterior mean for each parameter is indicated by a dot. . . . .	144
Figure 4.8	Individuals’ estimated weights from the perceptual model. <b>A.</b> Distributions of individual weights along various dimensions for L1 English listeners. <b>B.</b> Distributions of individual weights along various dimensions for L1 Mandarin listeners. Posterior means are represented by the dots, and the 89% CrIs are marked with error bars. . . . .	145
Figure 4.9	Post-stop F0 weights across perception and production for each L1 English and L1 Mandarin participant. Solid lines represent 100 regression lines fit with 100 posterior draws, to show direction and uncertainty in the correlation. . . . .	147
Figure 5.1	A possible response layout for the English non-CL and CL experiments. . . . .	163
Figure 5.2	Examples of visual displays used as cognitive load. <b>A.</b> Target-present grid, with the target (i.e., red square) in the fifth row and third column. <b>B.</b> Target-absent grid. . . . .	164

Figure 5.3	Illustration of trial procedure in the English cognitive load perception experiment. The overall procedure is similar to that of other perception experiments, except that the participant was additionally asked to pay attention to the grid during the playback of the syllable and search for a red square. The grid was displayed for 560 ms in all trials, and the participant had to indicate if they found a red square after they gave a response based on the audio stimulus. . . . .	165
Figure 5.4	Accuracy in the visual search task, based on “correct” target trials (see Section 5.2.5 for detail). A score was estimated for each listener in each block. The scores from the same listeners are connected by a line. <b>A.</b> Results for L1 English listeners who also participated in the non-CL version. <b>B.</b> Results for L1 English listeners who only did the CL version. <b>C.</b> Results for L1 Mandarin listeners. . . . .	168
Figure 5.5	Aggregated results on English stop-voicing categorization for the within-subject comparison, as a function of VOT, post-stop F0, tone, and CL conditions. <b>A.</b> With VOT on the $x$ -axis. <b>B.</b> With post-stop F0 on the $x$ -axis to highlight its effect on categorization. . . . .	171
Figure 5.6	Aggregated results on English stop-voicing categorization for the between-subject comparison, as a function of VOT, post-stop F0, tone, and CL conditions. <b>A.</b> With VOT on the $x$ -axis. <b>B.</b> With post-stop F0 on the $x$ -axis to highlight its effect on categorization. . . . .	172

Figure 5.7	Distributions of perceptual weights along various dimensions at the population level for <b>A.</b> the within-subject comparison and <b>B.</b> the between-subject comparison. Posterior means are represented by the dots. The 89% CrIs are marked by the inner error bars, and the 95% CrIs are marked by the outer error bars. The CrIs were created from 4,000 samples, each drawn from a normal distribution with the mean and standard deviation corresponding to those from a posterior sample output by the model. . . . .	173
Figure 5.8	Individuals' perceptual weights as estimated by the model. <b>A.</b> Distributions of individual weights along various dimensions under non-CL and CL conditions. <b>B.</b> Differences in cue weights along the same dimension across conditions. Posterior means are represented by the dots. The 89% CrIs are marked by the inner error bars, and the 95% CrIs are marked by the outer error bars. . . . .	174
Figure 5.9	Individuals' perceptual weights estimated by model. <b>A.</b> Distributions of individual weights along various dimensions for L1 English listeners under the non-CL condition. <b>B.</b> Distributions of individual weights along various dimensions for L1 English listeners under the CL condition. Posterior means are represented by the dots. The 89% CrIs are marked by inner error bars, and the 95% CrIs are marked by outer error bars. . . . .	175
Figure 5.10	Aggregated results on Mandarin stop-voicing categorization, as a function of VOT, post-stop F0, tone, and CL conditions. <b>A.</b> With VOT on the $x$ -axis. <b>B.</b> With post-stop F0 on the $x$ -axis to highlight its effect on categorization. . . . .	179

Figure 5.11	Distributions of perceptual weights along various dimensions at the population level. Posterior means are represented by the dots. The 89% CrIs are marked by the inner error bars, and the 95% CrIs are marked by the outer error bars. The CrIs were created from 4000 samples, each drawn from a normal distribution with the mean and standard deviation corresponding to those from a posterior sample output by the model. . . . .	181
Figure 5.12	Individuals' perceptual weights estimated by model. <b>A.</b> Distributions of individual weights along various dimensions for L1 Mandarin listeners under the non-CL condition. <b>B.</b> Distributions of individual weights along various dimensions for L1 Mandarin listeners under the CL condition. Posterior means are represented by the dots. The 89% CrIs are marked by inner error bars, and the 95% CrIs are marked by outer error bars. . . . .	182
Figure 6.1	Comparison of production and perceptual weights of post-stop F0 under different conditions. Note that the production and perceptual weights are on difference scales, so they are not directly comparable with each other. ES = L1 English speaker, MS = L1 Mandarin-L2 English speaker, EL = L1 English listeners, ML = L1 Mandarin-L2 English listener, Eng. = English task, and Man. = Mandarin task. . . . .	194
Figure 6.2	L1 Mandarin-L2 English bilinguals' production and perceptual weights for post-stop F0 in the English context as a function of individual summative language experience (SLE) scores. The lines represent linear regression results of post-stop F0 weight against SLE score. The shaded areas cover the 89% confidence interval of the regression. Each number represents a participant, and the same participant is represented by the same number in the two panels. . . . .	199

Figure 6.3	L1 Mandarin-L2 English bilinguals’ production and perceptual weights for post-stop F0 in the English context as a function of individual Bilingual Language Profile (BLP) scores. The lines represent linear regression results of post-stop F0 weight against BLP score. The shaded areas cover the 89% confidence interval of the regression. Each number represents a participant, and the same participant is represented by the same number in the two panels. . . . .	201
Figure F.1	Accuracy in the visual search task, based on both “correct” target and filler trials (see Section 5.2.5 for detail). A score was estimated for each listener in each block. The scores from the same listeners are connected by a line. <b>A.</b> Results for L1 English listeners who also participated in the non-CL version. <b>B.</b> Results for L1 English listeners who only did the CL version. <b>C.</b> Results for L1 Mandarin listeners. . . . .	290
Figure F.2	L1 English listeners’ guessing probabilities in each experimental condition. The dots mark the posterior means. The 89% CrIs are marked by the inner error bars, and the 95% CrIs are marked by the outer error bars. . . . .	291
Figure F.3	Guessing probabilities for L1 English listeners from both groups. The posterior means are marked by the dots. The 89% CrIs are marked by the inner error bars, and the 95% CrIs are marked by the outer error bars. . . . .	293
Figure F.4	Guessing probabilities for L1 Mandarin listeners from both groups. The posterior means are marked by the dots. The 89% CrIs are marked by the inner error bars, and the 95% CrIs are marked by the outer error bars. . . . .	295

# Acknowledgments

This dissertation wouldn't have been possible without many people. First and foremost is my supervisor, Kathleen Currie Hall, whose guidance, patience, meticulousness, support, insight, encouragement, and home-made ice cream were indispensable. The possibility of having a supervisor that possesses all these wonderful qualities is way too small, so I must have done something right in my previous life to deserve such an awesome supervisor!

The other members of my committee are also crucial in the development of this dissertation. The big-picture comments from Molly Babel were like a ring buoy when I was drowning in details. Her enthusiasm for linguistics also shone through her comments. And her ability to remember names and to suggest missing papers I should read is something that I can't hold a candle to... Márton Sóskuthy's attention to details, from pitfalls in my argumentation to data visualization, has made this dissertation less annoying to logicians and artists alike. His feedback on my statistical analyses and interpretations of model outputs has been tremendously useful in preventing me from embarrassing myself when presenting my research to the wider linguistic community.

I'd also like to acknowledge the funding from the 2021 Ministry of Science and Technology Taiwanese Overseas Pioneers Grants for PhD Candidates. Their fund helped pay for my shelter, and meat and mead for my last year as a graduate student.

Wisdom (and grant money) from other faculty members has benefited me greatly. I'm thoroughly dwarfed by Miikka Silfverberg's kindness, patience, dedication to teaching and research, and his tolerance when I fail to pronounce the geminated [k] in his name. His inability to say no meant that I could always seek

help from him. I was fortunate to have Doug Pulleyblank on my QP committee; his piercing insight and Socratic conversational style had always dragged me out of my phonetic comfort zone and prompted me to reexamine the issue from a more theoretical point of view. I should also thank Henry Davis for telling us how to distinguish ravens from crows.

I'd not have [Ph]inische[D] my PhD without peer support / pressure. My star PhD cohort—Khia Johnson, Gloria Mellesmoen, and Daniel Reisinger—was the best motivation for me to not fall behind. Daniel especially had successfully lured me into numerous movie evenings together. Other graduate students from UBC Linguistics turned out to be good not only at linguistics, but also at distracting me from my work. The top of the list includes (but not limited to) Ana Laura Arrieta-Zamudio, Yurika Aonuki, Alex Ayala, Weirui Chen, Sonja Frazier, Michael Fry (thank you for picking up from the airport when I first arrived in Vancouver), Mitchi Kamigaki-Baron, Aaditya Kulkarni, Zoe Lam, Suyuan Liu, Noah Luntzlara, Sander Nederveen, Bruce Oliver, Starr Sandoval, Rachel Soo, Oksana Tkachman, Kaili Vesik, Sijia Zhang, and Kate Zhou. I have spent too much good time with you guys. I'm also indebted to the following overly smart and overachieving people I met at UBC for their friendship: Boyan Beronov, Danny Bettencourt, Chien-Cho Chan, Pankaj Gupta, Rafael Haenel, Yuto Hirayama, Tomoharu Hirota, Brandon Hsu, Edwin Hsu, Lui Xia Lee, Edgar Liao, Nathaniel Lim, Millie Lou, Kyoko Matsumaga, Evan Koike, Hyunju Kwon, Mikko Paajanen, Nilan Saha, Nick Sanders, Camille Sung, Tarun Tummuru, Nila Utami, Christian Weilback, Oliver Yam, and Amitai Zand. Finally, *dankje* to my Dutch squad: Astrid Gilein, Gouming Martens, and Maxime Tulling.

Special thanks go to the past and current staff at UBC Linguistics. In particular, Murray Schellenberg spent many hours testing / re-testing / re-re-testing the room setup for my final defense. Claudia Chan helped me a lot with room booking at various points, and was always willing to chitchat with me. I also look forward to having more dim sum with you in the future!

最後的最後，要感謝一路扶持我的臺灣家人及朋友。謝謝家人無條件的包容及成全，說是三生有幸一點也不為過。謝謝我的老朋友，特別是蔡旻泓、陳弘哲、王季勤、黃敏雄、許雅雯、李薇（族繁不及備載），願意花時間聽我訴苦。但之後稱呼我的時候不要忘了加上博士的頭銜！

# Dedication

獻給天上的阿公  
及黑皮

To those who made  
my years in graduate school  
one hell of a ride.

# Chapter 1

## Introduction

### 1.1 Background

“Bill had a pill and a beer on the pier, then ate a peach on the beach while patting his pet bat.” Upon saying or hearing this tongue twister, most speakers of English can immediately recognize that it is playing with the /p/-/b/ distinction in the language. Under the disguise of this basic intuition, however, is a complex interplay of many acoustic dimensions. For instance, the main difference between /p/ and /b/ in English, and other stops in this voiceless-voiced series (i.e., /t/-/d/ and /k/-/g/) lies in the amount of time between the opening of the closure and the beginning of vocal fold vibration for the following vowel (i.e., voice onset time, or VOT), which is longer for /p/ than for /b/. VOT, nonetheless, is not the only difference between the two sounds; the pitch of the voice in the vowel is also slightly higher on average following /p/ than /b/. This phenomenon, where pitch or its physical correlate—fundamental frequency (F0)—shifts according to the voicing of a stop is referred to as post-stop F0 perturbation. Although both VOT and post-stop F0 are correlated with the stop voicing contrast in English, listeners do not ascribe equal importance to the two dimensions. In this case, VOT is listeners’ primary cue to the voicing distinction in English, and post-stop F0 plays the role of a secondary cue, which influences listeners’ decisions to a much smaller extent. In this work, it is this secondary cue—post-stop F0—that will take center stage. More specifically, one aim of the current work is to contribute to our understanding of how

speakers encode and how listeners decode post-stop F0 in contrasting stop voicing among first-language (L1) Mandarin and L1 English speakers. Additionally, using the post-stop F0 cue as a stepping stone, this study addresses production and perception in L1 and L2 (second language). Finally, it also examines whether and how L1 listeners of the two languages adapt their use of post-stop F0 when confronted with additional cognitive load.

Post-stop F0 represents an interesting dimension to look into because F0 plays different roles across English and Mandarin. In English, F0 is an important signifier for lexical stress (Bolinger, 1961; Fry, 1958; Lehiste, 1970) and intonation (Ladefoged and Johnson, 2014). The patterning between post-stop F0 and stop voicing is also well established: voiceless stops tend to raise the onset F0 of the following vowel while voiced stops tend to lower the F0 in production (Hanson, 2009; Hombert, 1978; Hombert et al., 1979; House and Fairbanks, 1953; Lea, 1973; Lehiste and Peterson, 1961; Ohde, 1984). In perception, studies have shown that English listeners pay attention to post-stop F0 when identifying stop categories, regardless of whether VOT, the primary cue for stop voicing, is ambiguous or not (Abramson and Lisker, 1985; Whalen et al., 1993). However, difference in F0 alone does not change the *lexical meaning* of word. For example, the syllable /bi/ pronounced with a high F0 still means the same type of insect as the identical syllable pronounced with a low F0.

Like English, Mandarin has a two-way laryngeal contrast in its stop series: this contrast is typically characterized as between voiceless aspirated (e.g., /p<sup>h</sup>/) and voiceless unaspirated (e.g., /p/). In spite of the nomenclature difference in classifying stop voicing across Mandarin and English, there are similarities in terms of phonetic implementation: the unaspirated / voiced class typically has a short-lag VOT (under 30 ms), and the aspirated / voiceless class a long-lag VOT (above 30 ms). Unlike English, however, difference in F0 alone is used to distinguish lexical items in Mandarin. That is, Mandarin has four lexical tones the distinction of which were primarily conveyed by changes in direction as well as height in F0 (see Section 2.1.1 for a detailed description of the tonal inventory of Mandarin). The use of F0 as the primary medium for the lexical tones in Mandarin (e.g., Gandour, 1978; Ohala, 1978) raises the question of whether F0 still functions as a cue for stop voicing in Mandarin, with respect to production and perception. This question

has been addressed in a number of studies (Chen, 2011; Guo, 2020; Howie, 1976; Luo, 2018; Xu and Xu, 2003; see Section 2.1.1 for a review of these studies). Concerning production, although all studies show that the presence or absence of aspiration in stops leads to different post-stop F0 profiles, they disagree with respect to whether aspiration leads to a higher or a lower post-stop F0. As for perception, the only systematic study, by Guo (2020), suggests that L1 Mandarin listeners use post-stop F0 as a cue for stop voicing in Mandarin. However, her experiment did not require the listener to track F0 for lexical tone while identifying the voicing feature. It is therefore still an open question as to whether L1 Mandarin listeners use post-stop F0 as a cue for voicing when it is simultaneously being used as the primary cue for tone. These yet-to-be-solved issues regarding the production and perception of post-stop F0 in Mandarin are addressed in Chapter 2 and Chapter 3.

Another angle from which the use of post-stop F0 is examined in this work is bilingualism. Both flexibility and constraint have been observed in the linguistic systems of bilingual speakers. For examples, Amengual (2021) examined the VOT of the English and Japanese /k/ in the production from L1 English-L2 Japanese and L1 Japanese-L2 English bilinguals, and found that both groups of speakers produced language-specific VOT patterns for each language. However, it is also well established that L1 can exert a strong influence on L2, especially when L2 is acquired later in life. One of the most famous examples is the notorious difficulty of the English /r/-/l/ distinction for L1 Japanese listeners, which can be attributed to the fact that the English contrast relies primarily on a difference in third formant values, whereas L1 Japanese listeners distinguish the categories primarily based on second formant values (Iverson et al., 2003; Miyawaki et al., 1975). The relatively large population of L1 Mandarin-L2 (second language) English speakers in Canada provides an opportunity to investigate whether bilingual speakers of tonal and non-tonal languages use post-stop F0 similarly across different language contexts. Comparing L1 Mandarin-L2 English bilinguals' use of post-stop F0 across the two languages allows us to evaluate whether post-stop F0 is a dimension that can be controlled by speakers, which also has implications for the mechanism behind post-stop F0 perturbation. In addition, by contrasting the bilinguals' use of post-stop F0 in the English context with that of L1 English speakers, the potential influence of L1 Mandarin on L2 English in this acoustic dimension can be assessed.

These topics of adaptability of post-stop F0 in bilingual speakers are taken up in Chapter 3 and Chapter 4.

Finally, given that listening conditions in everyday life are filled with environmental noise that competes for listeners' selective attention, it is important to understand whether and how the use of post-stop F0 in perception is impacted in adverse listening conditions. This question motivates the experimental study reported in Chapter 5, which explores the use of post-stop F0 in non-ideal listening conditions simulated by introducing additional cognitive load.

## **1.2 The Current Study**

The current work is an investigation of the use of post-stop F0 in production and perception by L1 Mandarin-L2 English bilinguals and L1 English listeners, focusing on three different aspects. The first aspect concerns L1 Mandarin-L2 English bilinguals' flexibility in the use of post-stop F0 in Mandarin and English contexts. The second aspect focuses on the L1 influence on the use of post-stop F0 in L1 Mandarin-L2 English bilinguals, as compared to L1 English speakers. The third aspect targets whether and how cognitive load changes the perceptual use of post-stop F0 in L1 Mandarin-L2 English bilinguals as well as L1 English listeners. When this research project was initially conceived, all the proposed perception experiments were in the format of the visual world eye-tracking paradigm (Cooper, 1974; Tanenhaus et al., 1979, 1995). The eye-tracking technique as a tool to investigate cognitive processing is desirable in this research context because past phonological and phonetic processing research using it has demonstrated that spoken word recognition exhibits graded sensitivity to within-category VOT (Clayards et al., 2008; McMurray et al., 2008, 2002, 2009). That is, when the proportion of fixations is plotted against manipulated VOT values, a more linear trend, as opposed to an S-shaped curve based on the force choice task, is observed. Perception experiments coupled with eye-tracking should therefore be more revealing about how various acoustic dimensions are processed by the listener and how this process unfolds over time. And then came COVID-19. The university banned all in-person research activity, so the original eye-tracking experiments were abandoned and replaced with online mouse-tracking experiments (as described in Section 3.4.3) as

mouse trajectories have been claimed to be able to reflect decision-making processes (Grage et al., 2019; Kieslich et al., 2019, 2020; Maldonado et al., 2018; Schoemann et al., 2019, 2021; Spivey et al., 2005; Wulff et al., 2019). However, preliminary analyses on mouse-tracking data showed no discernible patterns, with an excessive amount of noise. In fact, a later study has showed that mouse-tracking might not have enough resolution to tap into within-category sensitivity revealed in previous eye-tracking studies (Stoeber, 2019). Hence, although mouse-tracking tasks were used, only click response data (i.e., accuracy and reaction time) was analyzed and included in this work.

The empirical focus of the work is the stop voicing contrast in Mandarin and English. Mandarin and English share some of the same cues in implementing the contrast; however, the specific way these cues are used and the distinct phonological functions of one of the cues—F0—vary by language. Along with comparing cross-linguistic differences in the use of post-stop F0, another focus of the current work is the extent to which individuals with the same language background differ in their use of the post-stop F0 cue, as well as how individual listeners' perceptual weighting of this cue relates to the way they use it in their own production. After examining the use of post-stop F0 in presumably ideal conditions, the final experimental study explores whether and how listeners shift their use of the cue to adapt to listening conditions laden with additional cognitive load.

The specific research questions and the chapters that address each research question are listed below:

- **RQ1:** Is post-stop F0 in L1 Mandarin speakers' production correlated with the phonological voicing of stops in the language?
  - Chapter 2 addresses this question via a corpus study that analyzed broadcast news speech from L1 Mandarin speakers.
  - Chapter 3 approaches this question with a production experiment where L1 Mandarin-L2 English bilinguals produced Mandarin words whose initial consonants typified the voicing contrast.
- **RQ2:** How do L1 Mandarin-L2 English bilinguals weight post-stop F0 in their *production* of Mandarin and English stops? Does the language context

play a role in their use of post-stop F0 in production?

– Chapter 3 answers these questions with a pair of Mandarin and English production experiments that followed the same procedure.

- **RQ3:** Relatedly, how do L1 Mandarin-L2 English bilinguals weight post-stop F0 in their *perception* of Mandarin and English stops? Again, does the language context modulate the use of post-stop F0 in perception?

– Chapter 3 tackles these questions through a pair of Mandarin and English perception experiments that shared the same critical stimuli and experimental paradigm.

- **RQ4:** How does L1 Mandarin-L2 English bilinguals' production and perceptual use of post-stop F0 as a cue for the stop voicing in English compare to that of L1 English speakers?

– Chapter 4 investigates the question by comparing production and perception data from the bilingual and L1 English listeners.

- **RQ5:** On both the population and individual levels, is the production weight of post-stop F0 predictive of the perceptual weight of the same cue?

– Chapter 3 and Chapter 4 deal with this question by carrying out a series of correlational analyses.

- **RQ6:** How does cognitive load affect the use of post-stop F0 in identifying stop voicing in Mandarin and English?

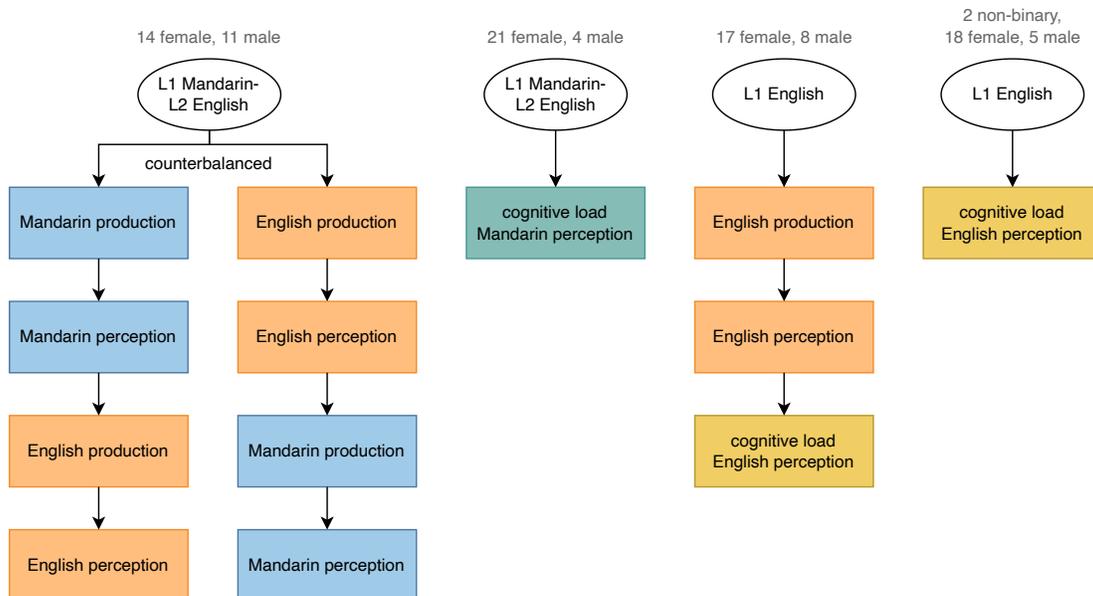
– Chapter 5 attempts to provide an answer to this question by the perceptual use of post-stop F0 across load and non-load conditions by L1 Mandarin-L2 English and L1 English listeners.

Similar strategies are used to quantify speakers' and listeners' use of the post-stop F0 cue across the various experimental studies in the current work. To measure speakers' use of post-stop F0 in production, participants read a series of stop-initial monosyllabic words. The acoustic measurements for post-stop F0 are then taken

from the target consonants, and the distributions of values on this dimension are compared across the segments of the contrast. The degree to which these distributions are separated for the post-stop F0 cue is used as a metric of speakers' use of it in production (Clayards, 2018). The perception experiments involve forced-choice identification tasks, in which listeners hear stop-initial target stimuli (and filler stimuli with non-stop initials) that have been systematically manipulated along the relevant acoustic dimensions. Listeners are tasked to determine which sound they hear for each stimulus (e.g., /p/, /b/, /m/, or /n/), and listeners' perceptual weight for post-stop F0 is quantified on the basis of the corresponding coefficient from logistic regression models indicating the extent to which their responses can be predicted using the value of the cue (Morrison and Kondaurova, 2009).

As hinted at above, this work adopts corpus-based and experimental approaches to address the proposed research questions. An overview of the experiments involved is given in Figure 1.1. The parallel studies across languages and modalities result in a database that can be used to explore speakers' and listeners' use of the post-stop F0 cue, and the corresponding analysis can inform theories about the origin and function of the post-stop F0 perturbation (Halle and Stevens, 1971; Hombert et al., 1979; Kingston and Diehl, 1994; Kingston et al., 2008; Kohler, 1984; Ladefoged, 1967; Löfqvist, 1975; Ohala and Ohala, 1972; Slis, 1970). One major question running through the parallel production-perception experiments presented here is the link between production and perception. Such an investigation is therefore relevant for testing predictions of models positing a strong production-perception tie (e.g., Liberman and Mattingly's [1985] Motor Theory and Fowler's [1986] Direct Realism).

In sum, this dissertation presents a close examination of the use of post-stop F0 in defining parallel sound contrasts across languages and modalities. Data from L1 English and L1 Mandarin populations puts forth a testing ground to compare different roles of post-stop F0 in contrasting stop voicing cross-linguistically. The analyses also foreground the extent to which individuals vary in the use of post-stop F0, as well as how the particular use of F0 in L1 influences the usage patterns in an L2. More broadly, this work attempts to explore the questions of how speakers and listeners take advantage of acoustic dimensions that serve multiple phonological functions, to what degree the use of a cue is dictated by the ambient language



**Figure 1.1:** An overview of the experiments conducted in this work, together with the composition of participants in each experiment. The same target audio stimuli are used in all perception experiments.

environment, and listeners’ strategies to adapt cue weights when confronted with adverse listening conditions.

### 1.3 Organization of Chapters

Chapter 2 begins with a review of experimental investigation of the post-stop F0 perturbation in Mandarin and then presents a corpus study of the same phenomenon with data from Mandarin broadcast news speech. The corpus study aims to understand the conflicting results from previous studies through a reanalysis using two different normalization strategies: F0 standardization and F0 scaling. The results based on the standardization method suggest that vowel-onset F0 after unaspirated stops was higher than that after aspirated stops or after sonorants. The results from the scaling method, however, indicate that F0 after aspirated stops dropped faster than that after unaspirated stops, and that F0 after unaspirated stops in turn fell more quickly than that after sonorants.

Chapter 3 provides a systematic experimental investigation of L1 Mandarin-L2 English bilinguals' use of post-stop F0 in their first and second languages. This case study serves two purposes: (i) to revisit the topic of the role of post-stop F0 in Mandarin, as previous studies have shown conflicting outcomes, and (ii) to explore the L1-L2 flexibility with respect to post-stop F0. Using a production and a perception task, cue weights are calculated for the post-stop F0 dimension across modalities (i.e., production and perception) as well as languages (i.e., English and Mandarin). To foreshadow the results, the bilinguals produced a higher post-stop F0 after an aspirated stop than after an unaspirated stop in both Mandarin and English. They also used post-stop F0 as a perceptual cue for stop voicing in either a Mandarin or an English context, with a higher post-stop F0 leading to more aspirated responses. Furthermore, the language context did modulate post-stop F0 weights: across both production and perception, post-stop F0 was weighted more in the English context than in the Mandarin context.

While Chapter 3 focuses on L1 Mandarin-L2 English bilinguals' cross-linguistic performance, Chapter 4 addresses how their L2 English performance compares to that of L1 English speakers. With the same production and perception experimental paradigm as in Chapter 3, L1 Mandarin-L2 English bilinguals' usage patterns of post-stop F0 in their L2 English are compared with those of L1 English speakers. The production data shows that the bilinguals and L1 English speakers used post-stop F0 similarly: both groups had comparable production weights. The perception data, on the other hand, reveals that, although both groups used post-stop F0 as a cue, the bilinguals relied on post-stop F0 to a lesser degree than L1 English listeners. This comparison therefore provides a basis for a more nuanced understanding of the degree to which L1 Mandarin-L2 English bilinguals' use of post-stop F0 in the English stop contrast is biased by cue use in their native language.

Chapter 5 turns to whether and how additional cognitive load modulates the perceptual use of the post-stop F0 cue, in addition to VOT, in L1 English and L1 Mandarin listeners. Specifically, this study focuses on how the two groups of listeners adapt their cue-weighting strategies when instructed to perform a secondary task that is concurrent to the word-identification task. To anticipate the results, the analyses suggest that cognitive load did not shift the average perceptual weight of

VOT or post-stop F0 for either listener group. Instead, cognitive load enlarged the variance of the two cues, at least for L1 English listeners.

Finally, Chapter 6 provides a summary of main findings, responses to the research questions, a general cross-chapter discussion, and a concluding remark.

## **Chapter 2**

# **A Corpus Study on Vowel-Onset F0 in Mandarin**

### **2.1 Introduction**

It has long been observed that, in many languages, the F0 at the beginning of the vowel following a stop consonant can signal differences in the stop's phonological voicing. This phenomenon is also commonly labeled as pitch skip, obstruent intrinsic F0, co-intrinsic pitch, onset F0 perturbation, or post-stop F0 perturbation. Over the past 50 years, vowel-onset F0 has received considerable attention, and methods and interpretations of the results also evolved along the way (as reviewed in Abramson and Whalen [2017] and references therein). Although the detailed findings vary across studies, it is generally agreed that, at least in languages with a two-way laryngeal contrast (e.g., Spanish and English), F0 at vowel onset is significantly higher following phonologically voiceless stops than following phonologically voiced ones, regardless of the presence of actual vocal fold vibration (Dmitrieva et al., 2015; Hanson, 2009; Kingston and Diehl, 1994).

Following this line of research, the research question this study asks is whether and how the phonological status of the initial consonant has an impact on vowel-onset F0 in Mandarin. In answering the research question, this study serves two functions: (i) it summarizes past studies on the influence of phonological voicing on vowel-onset F0 in Mandarin, which lies at heart of this dissertation, and (ii) it

attempts to understand some of the contradictory findings, as will be revealed in Section 2.1.1, by means of a reanalysis of data from a speech corpus, using two different F0 normalization strategies. The comparison between the two normalization strategies then further informs the F0 normalization method adopted in subsequent chapters.

As discussed in Chapter 1, Mandarin serves as the model language because of its use of F0 as the primary cue for its lexical tone (see Section 2.1.1 below), which raises the point of a potential dual role of F0 in signaling both lexical tone and phonological voicing in the language. This study resorts to a corpus-based approach, complementing previous experimental investigations (as will be reviewed in Section 2.1.1) on similar research questions. While experimental studies thrive in controlling potential confounds and establishing causal relationships, the laboratory settings where experimental studies are typically conducted might also impact language use. Corpus studies with data captured outside laboratories have a better chance to reflect language use in a more natural setting. Given that previous systematic research is experimental in nature, this corpus study sheds light on the research question from a different perspective.

Across both experimental and corpus studies, it should also be noted that various research procedures and measurements were involved in previous cross-linguistic studies on vowel-onset F0. Some studies used mean F0 over the entire vowel (e.g., House and Fairbanks, 1953), some measured the maximum F0 in the vowel (e.g., Lehiste and Peterson, 1961), and some tracked F0 throughout the vowel (e.g., Guo, 2020). In terms of stimuli, both real words (e.g., Kirby, 2018) and nonwords (e.g., Lea, 1973) were used, and both monosyllabic words (e.g., Luo, 2018) and disyllabic words (e.g., Xu and Xu, 2003) were employed. As for the presentation of the stimuli, some studies presented the stimuli in isolation (e.g., Lehiste and Peterson, 1961) while some embedded the stimuli into a carrier phrase (e.g., Kirby and Ladd, 2016). These methodological differences are especially important when discussing studies on vowel-onset F0 perturbation in Mandarin, as conflicting results have been reported, and the methodological differences across these studies might be a contributing factor. Therefore, when reviewing these studies on Mandarin in the next section, I put a special focus on their methodology.

### 2.1.1 Previous Studies on Vowel-Onset F0 in Mandarin

The existing literature on the vowel-onset F0 perturbation effect in Mandarin depicts a mixed picture, with conflicting results across studies. These studies are summarized in Table 2.1 and described below.

Mandarin has six stops coming in unaspirated-aspirated pairs: /p/-/p<sup>h</sup>/, /t/-/t<sup>h</sup>/, and /k/-/k<sup>h</sup>/, and has four lexical tones: high-level ˩ (Tone 1), mid-rising ˨˨˨˨˩ (Tone 2), low-dipping ˨˨˨˨˩˨˩ (Tone 3), and high-falling ˨˨˨˨˩˨˩˨˩˨˩ (Tone 4). In an early study by Howie (1976), which examined the differences in the pitch pattern of each tone in response to various types of segments that preceded the vowel, one male speaker of Mainland Mandarin produced various monosyllabic words, which exemplified different onsets (including /p, p<sup>h</sup>, t, t<sup>h</sup>, k, k<sup>h</sup>/), rimes, and tones, embedded in a carrier phrase. F0 was measured manually using the distance between the first and second harmonics over the voiced part of the syllable. As only one speaker was recorded, no F0 normalization was performed. Neither was any statistical analysis conducted. However, visual inspection of the resulting F0 trajectories by the current author suggests that, generally speaking, the vowel-onset F0 following an aspirated stop tended to be higher than that following an unaspirated stop, while the onset F0 following an unaspirated stop was in turn higher than that following a sonorant.

Xu and Xu (2003) recruited seven female speakers, aged between 22 and 30 years, to produce disyllabic words with the target syllables /t<sup>h</sup>a/ and /ta/ (in all four tones) in either the first or second position. These stimuli were presented to the speakers in two carrier phrases that aimed to further vary the tonal context before the stimulus. F0 was extracted over the entire vowel using automatic vocal cycle detection and manual rectification (Xu, 1997), and the F0 trajectories were smoothed individually with a three-point median filter. As each stimulus was repeated five times, the five repetitions of each syllable by each speaker were averaged to remove random variation, and subsequent statistical analyses were performed on these averaged values. No F0 normalization was used in the analyses. Their results indicated that, across the four lexical tones, /t<sup>h</sup>/ led to a *lower* vowel-onset F0 than /t/. This finding is different from Howie (1976) and from that reported in other languages (e.g., English [Hanson, 2009], French [Kirby and Ladd,

2016], Japanese [Gao and Arai, 2018], Spanish [Dmitrieva et al., 2015]). However, the magnitude of vowel-onset F0 difference between /t<sup>h</sup>/ and /t/ depended on tone: Tone 2 and Tone 3 supported a greater F0 difference than Tone 1 and Tone 4.

Chen (2011) investigated the effect of speech rate on the impact of aspiration on F0 using a delayed shadowing paradigm, where the participant had to repeat what they heard with the same speaking rate after a 500-ms delay. Twenty speakers of Taiwan Mandarin (10 female and 10 male, aged between 20 to 30 years) imitated a female model speaker who produced monosyllables inserted into a carrier phrase at a fast and a slow rate. These tonal monosyllables contained onsets that typified the six stops in Mandarin and vowels /i, a, u, ə/ in all four lexical tones. Praat (Boersma and Weenink, 2021) was used to measure F0, but no algorithmic detail was provided. F0 was also not normalized prior to the analyses; instead, gender was included as a factor in the model to account for the associated F0 variation. Her results stood in contrast to those reported in Xu and Xu (2003), with vowel-onset F0 being higher after aspirated stops, as opposed to unaspirated ones, in all tonal contexts. However, tone and gender seemed to condition vowel-onset F0 as well: for male speakers, the F0 difference was larger in Tone 1 and Tone 4 than in Tone 2 and Tone 3, but for female speakers, the opposite was observed.

Using tonal monosyllables with onsets /p, p<sup>h</sup>, t, t<sup>h</sup>, k, k<sup>h</sup>, m, n, l/, Luo (2018) similarly examined the influence of onset voicing on the F0 trajectory of the following vowel. F0 was measured using Praat (Boersma and Weenink, 2021) from the production data of 15 female speakers (aged between 19 and 33 years). F0 data was standardized (i.e., z-transformed) within each speaker before being fed into the statistical models. Her results indicated that vowel-onset F0 was highest after an aspirated stop and lowest after a sonorant, with F0 after an unaspirated stop in-between. Her finding is therefore in line with Howie (1976) and Chen (2011), but different from Xu and Xu (2003). Lexical tone again was found to have an effect: Tone 3 induced a greater F0 difference between aspirated and unaspirated stops than Tone 1 or Tone 4, but post-stop F0 difference was not observed in the context of Tone 2.

Finally, 25 speakers (15 female, 10 male, aged between 19 and 46 years) participated in Guo's (2020) study, in which they read aloud monosyllabic words (with onsets /p, p<sup>h</sup>, t, t<sup>h</sup>, w/, vowels /a, u/, and all four tones) incorporated into a car-

rier phrase. Praat (Boersma and Weenink, 2021) was used to estimate F0 over the length of the vowels, and F0 values were standardized ( $z$ -transformed) within each speaker. Using linear mixed-effects models, she found that the direction of vowel-onset F0 perturbation hinged on tone: the vowel-onset F0 following aspirated stops was significantly *higher* than that following unaspirated stops in Tone 1 and Tone 4, but was significantly *lower* than that following unaspirated stops in Tone 2 and Tone 3. Her results are thus “in-between” what has been reported, aligning with Howie (1976), Chen (2011), and Luo (2018) when it comes to Tone 1 and Tone 4, but in accordance with Xu and Xu (2003) when Tone 2 and Tone 3 are concerned. The vowel-onset F0 difference between aspirated/unaspirated stops and sonorants was also influenced by tonal environments. Specifically, the F0 after aspirated stops was higher than that after sonorants in Tone 1 and Tone 4, but not in Tone 2 or Tone 3. As for the vowel-onset F0 difference between unaspirated stops and sonorants, unaspirated stops consistently led to a higher vowel-onset F0 than sonorants in all four tonal contexts.

In sum, three studies showed that, overall, vowel-onset F0 after an aspirated stop was higher than that after an unaspirated stop (Chen, 2011; Howie, 1976; Luo, 2018), one study indicated the opposite pattern (Xu and Xu, 2003), and one study suggested that the pattern was coupled with the lexical tones (Guo, 2020). These diverse findings, as well as other cross-linguistic studies on vowel-onset F0 perturbation, have prompted a search of mechanisms that aim to understand the cause of the F0 perturbation effect. We turn to these mechanisms in the next section.

**Table 2.1:** Summary of previous experimental studies on vowel-onset F0 perturbation in Mandarin.

	<b>Howie (1976)</b>	<b>Xu and Xu (2003)</b>	<b>Chen (2011)</b>	<b>Luo (2018)</b>	<b>Guo (2020)</b>
<b>Participants</b>	1 male (age not provided); from Beijing and lived in the US when recorded.	7 female (age: 22–30 years); had lived in the US for about two years.	10 female and 10 male (age: 20–30 years); lived in Taiwan.	15 female (age: 19–33 years); 7 from Beijing and 8 from Shenzhen.	15 female and 10 male (age: 19–46 years); had lived in the US for about one year.
<b>Mandarin variety</b>	Mainland	Mainland	Taiwan	Mainland	Mainland
<b>Stimuli</b>	136 monosyllables with <b>onsets:</b> /p, p <sup>h</sup> , t, t <sup>h</sup> , k, k <sup>h</sup> /, among others; <b>rimes:</b> various types including high, non-high, front, back, rounded, and unrounded vowels; <b>tones:</b> all four lexical tones.	Disyllabic words consisting of syllables with <b>onsets:</b> /t, t <sup>h</sup> , m, ʃ/; <b>rime:</b> /a/; <b>tones:</b> all four tones (/p <sup>h</sup> a2/ in place of /t <sup>h</sup> a2/ due to lexical gap). The target syllables (/ta/, /t <sup>h</sup> a/) were either in the first position or in the second position, with all possible tonal contexts.	Monosyllables with <b>onsets:</b> /p, p <sup>h</sup> , t, t <sup>h</sup> , k, k <sup>h</sup> /; <b>rimes:</b> /i, a, u, ə/; <b>tones:</b> all four tones. Note that /ki/ and /k <sup>h</sup> i/, which are phonological gaps in Mandarin, were also included.	Monosyllables with <b>onsets:</b> /p, p <sup>h</sup> , t, t <sup>h</sup> , k, k <sup>h</sup> , m, n, l/; <b>rimes:</b> /i, a, u, ə/; <b>tones:</b> all four tones. Note that /ki/ and /k <sup>h</sup> i/, which are phonological gaps in Mandarin, were excluded.	Monosyllables with <b>onsets:</b> /p, p <sup>h</sup> , t, t <sup>h</sup> , w/; <b>rimes:</b> /a, u/; <b>tones:</b> all four tones.
<b>Carrier phrase</b>	<i>zhe4ge ___ zi4 shi4 lao3li3 xie3 de.</i> ‘This word ___ was written by Laoli.’	<i>wo3 lai2 shuo1 ___ zhe4ge ci2.</i> ‘I say the word ___.’ and <i>wo3 lai2 zhao3 ___ zhe4ge ci2.</i> ‘I look for the word ___.’	<i>wo3 nian4 ___ zhe4ge zi4.</i> ‘I read aloud the word ___.’	<i>wo3 shuo1 ___ zi4 san3 ci4.</i> ‘I say the word ___ three times.’	<i>qiang3 shuo1 ___ yi2 ci4.</i> ‘Please say ___ once.’

**Table 2.1 continued:** Summary of previous experimental studies on vowel-onset F0 perturbation in Mandarin.

	<b>Howie (1976)</b>	<b>Xu and Xu (2003)</b>	<b>Chen (2011)</b>	<b>Luo (2018)</b>	<b>Guo (2020)</b>
<b>Procedure</b>	Once in a randomized order	Five repetitions randomly	Delayed (500 ms) shadowing paradigm: repeated what was heard with the same speaking rate. Each token was repeated three times.	Three repetitions randomly	3 repetitions/block $\times$ 2 blocks = 6 repetitions in total
<b>F0 measurement</b>	<ul style="list-style-type: none"> <li>- Manual</li> <li>- Estimated using the distance between the first and second harmonics</li> <li>- Over the duration of the rime at equidistant points</li> </ul>	<ul style="list-style-type: none"> <li>- Automatic + manual rectification</li> <li>- Estimated using vocal cycle detection and smoothed with a three-point median filter (Xu, 1997)</li> <li>- Over the duration of the rime</li> <li>- Five repetitions by each speaker were averaged</li> </ul>	<ul style="list-style-type: none"> <li>- Automatic</li> <li>- Measured using Praat (Boersma and Weenink, 2021) (no algorithmic details provided)</li> <li>- (i) duration-neutralized F0 over the vowel, (ii) F0 over the first 100 ms of the vowel, (iii) vowel-onset F0</li> </ul>	<ul style="list-style-type: none"> <li>- Automatic</li> <li>- Measured with Praat (Boersma and Weenink, 2021) (500-Hz pitch ceiling at every 5 ms)</li> <li>- (i) F0 values within the first 30 ms of the vowel, (ii) time-neutralized F0 values at 11 equidistant points along the entire vowel</li> </ul>	<ul style="list-style-type: none"> <li>- Automatic</li> <li>- Measured using Praat (Boersma and Weenink, 2021) (no further details given)</li> <li>- (i) taken every 8 ms within the first 64 ms of the vowel, (ii) time-neutralized F0 values at 20 equidistant points over the vowel</li> </ul>
<b>F0-normalization strategy</b>	None (only one speaker)	None (all speakers being female)	None (gender was included as a predictor in the analysis)	standardization (z-transformation) within each speaker	standardization (z-transformation) within each speaker
<b>Statistical analysis</b>	None	repeated-measures ANOVA	repeated-measures ANOVA	repeated-measures ANOVA	linear mixed-effects model (with a by-speaker random structure)

**Table 2.1 continued:** Summary of previous experimental studies on vowel-onset F0 perturbation in Mandarin.

	<b>Howie (1976)</b>	<b>Xu and Xu (2003)</b>	<b>Chen (2011)</b>	<b>Luo (2018)</b>	<b>Guo (2020)</b>
<b>Overall results</b>	<ul style="list-style-type: none"> <li>- F0 after aspirated higher than F0 after unaspirated</li> <li>- F0 after unaspirated higher than F0 after sonorant</li> </ul>	<ul style="list-style-type: none"> <li>- F0 after aspirated <i>lower</i> than F0 after unaspirated</li> <li>- Magnitude of F0 difference varies by tone</li> </ul>	<ul style="list-style-type: none"> <li>- F0 after aspirated higher than F0 after unaspirated</li> <li>- Magnitude of F0 difference varies by tone</li> <li>- Tone effects vary by gender</li> </ul>	<ul style="list-style-type: none"> <li>- F0 after aspirated higher than F0 after unaspirated</li> <li>- F0 after unaspirated higher than F0 after sonorant</li> <li>- Magnitude of F0 difference varies by tone</li> </ul>	<ul style="list-style-type: none"> <li>- Direction of F0 perturbation varies by tone</li> <li>- F0 after aspirated higher than F0 after unaspirated in Tone 1 and Tone 4</li> <li>- F0 after aspirated <i>lower</i> than F0 after unaspirated in Tone 2 and Tone 3</li> <li>- F0 after unaspirated higher than F0 after sonorant across all tones</li> </ul>

### 2.1.2 Mechanisms behind Post-Stop F0 Perturbation

The mechanisms that have been evoked to account for the split of vowel-onset F0 according to voicing can be categorized into three camps: automatic, controlled, and hybrid. *Automatic* processes refer to accounts based on the articulatory or aerodynamic properties of the production of stops, while *controlled* processes are understood as mechanisms involving active, though subconscious, control on the part of speakers, which is language specific and forms part of a speaker's phonological knowledge. *Hybrid* processes, as the name suggests, take ideas from the two former processes. In what follows, I first review explanations based on automatic processes before turning to those based on controlled and hybrid processes.

Various articulatory and aerodynamic mechanisms have been suggested to elucidate the cause of post-stop F0 perturbation (e.g., Halle and Stevens, 1971; Hombert et al., 1979; Kohler, 1984; Ladefoged, 1967; Löfqvist, 1975; Ohala and Ohala, 1972; Slis, 1970). One such articulatory mechanism relies on *larynx height*. Specifically, F0 lowering after voiced compared to voiceless stops may be attributed to the lowering of the larynx during closure (Ewan, 1976). Larynx lowering, which helps to sustain voicing during closure, may cause the cricoid cartilage to rotate and consequently result in a lower F0 (Honda et al., 1999). Another articulatory mechanism is the combination of *cricothyroid muscles* and *vocal fold tension*. That is, during the closure of voiceless stops, tension in the cricothyroid musculature leads to stiffening of the vocal folds, which prevents phonation during voiceless stops (Halle and Stevens, 1971; Löfqvist et al., 1989). Such tension in the vocal folds is attenuated during the closure of voiced stops. This effect of different tension levels of the vocal folds during closure spreads to the following vowel, with tense vocal folds in a voiceless stop leading to a higher F0, and slack vocal folds in a voiced stop to a lower F0.

On the aerodynamic side, *trans-glottal pressure*, which refers to the pressure difference between the subglottal pressure and the oral pressure, has been implicated. In producing a stop, oral pressure builds up during closure and drops upon the release of the stop. However, the oral pressure accumulates faster for voiceless stops than for voiced stops thanks to the open glottis during voiceless stops (Ladefoged, 1971). With high oral pressure, voiceless stops induce faster trans-

glottal airflow at vowel onset and therefore raise the F0. On the other hand, lower oral pressure associated with voiced stops does not induce a fast trans-glottal airflow, relative to voiceless stops, so a lower F0 is observed at vowel onset (Kohler, 1984; Ladefoged, 1967). Difference in *subglottal pressure* has also been examined. A constant subglottal pressure is usually maintained during closure for all stops (Ladefoged, 1967; Löfqvist, 1975; Ohala and Ohala, 1972; Slis, 1970). Upon the release of an aspirated stop, a high rate of airflow runs through the glottis, and subglottal pressure decreases markedly during the aspiration period. In contrast, upon the release of an unaspirated stop, subglottal pressure decreases fairly gradually due to relatively little air flowing out of the subglottal area. Lower subglottal pressure is therefore generally observed after an aspirated stop than an unaspirated one (Ladefoged, 1963, 1974; Ohala, 1978; Ohala and Ohala, 1972). Since lower subglottal pressure is correlated with a lower F0, aspirated stops should give rise to a lower vowel-onset F0 than unaspirated ones. In sum, all of these properties are thought to be *automatic*, and as such, would be predicted to be the same across all languages.

It should be noted that three out of the four articulatory and aerodynamic mechanisms reviewed above concern phonetically voiceless and voiced stops, and all of these three processes predict the same effect: F0 should be lower after voiced stops and higher after voiceless stops. Only the last of them—the subglottal pressure mechanism—explicitly addresses the differences between voiceless aspirated and voiceless unaspirated stops. Given that the implementation of the phonological voicing contrast in Mandarin hinges on aspiration instead of phonetic voicing, the subglottal pressure mechanism would predict aspirated stops to be correlated with a lower post-stop F0 and unaspirated ones to be correlated with a higher post-stop F0. This is incompatible with some of the empirical findings reported above (Chen, 2011; Guo, 2020; Howie, 1976; Luo, 2018). Additionally, the predictions made by the first three accounts are not consistent with some of the cross-linguistic findings. For example, Kingston and Diehl (1994) show that F0 is raised after voiceless unaspirated stops in languages such as French, where this is the canonical phonetic implementation of [–voice] stops, but lowered next to such stops in languages such as English, where this is the implementation of [+voice] stops. These findings are hard to reconcile with the claim made by automatic approaches

that intended articulation automatically influences F0 as the F0 difference does not appear to depend in any consistent way on the other phonetic properties of the stop. These findings thus motivate the proposal that views F0 perturbation as a controlled process (Kingston, 2007; Kingston and Diehl, 1994).

The perturbation-as-controlled mechanism posits that the phonological status of the consonant carries more weight in determining the onset F0 patterns than do its phonetic properties. Crucially, these onset F0 patterns are language-specific and part of subconscious phonological knowledge, with the aim being to enhance the perceptual salience of the contrastive laryngeal features. Aside from Kingston and Diehl's (1994) study on English and French, Dmitrieva et al.'s (2015) finding that English short-lag and lead voice onset time (VOT) stops, which are subphonemic variants of the same [+voice] phonological category, do not differ in terms of onset F0, and that the initial short-lag stops in English, which are [+voice], are associated with a lower vowel-onset F0 than the initial short-lag stops in Spanish, which are [–voice], also provides support for the controlled mechanism.

Finally, it should be noted that the automatic and controlled views of F0 perturbation are not irreconcilable—more research has begun to converge on a hybrid approach that combines the ideas expressed in both mechanisms. For example, Hoole and Honda (2011) propose that the cricothyroid activity patterns, with their origin in the articulatory properties of voicing production, can be deliberately exaggerated by speakers to enhance the perceptual distinctiveness of the voicing contrast.

Implicitly, all the accounts described above focus on F0 variation due to phonological/phonetic voicing *within* a single speaker. However, in a community with multiple speakers, *between-speaker* F0 variation can originate from a host of factors, such as biological sex, constructed gender, age, and other physiological and social elements. In the current study, with its focus on F0 variation rooted in phonology, it is important to try to minimize the between-speaker variation in F0 due to physiological and social factors, so the F0 patterns uncovered can be compared across speakers. This is when F0 normalization strategies enter the picture, which will be reviewed immediately below.

### 2.1.3 F0 Normalization Strategies

Disner (1980), Thomas (2002), and Thomas (2011) list four general goals of a normalization technique, though not all techniques meet all of these goals:

1. Eliminating variation caused by physiological differences among speakers (e.g., differences in vocal fold lengths);
2. Preserving sociolinguistic/dialectal/cross-linguistic differences in the acoustic dimension of interest;
3. Preserving phonological distinctions along the dimension of interest;
4. Modeling the cognitive processes that allow human listeners to normalize acoustic dimensions uttered by different speakers.

In the present research context, the first and the third goals are the main driver for adopting an F0 normalization method. In particular, given that the speech data analyzed in this study comes from a corpus of 20 speakers, it is necessary to remove the effects of talker differences due to physiological differences before drawing a conclusion that characterizes the community as a whole. In addition, it is important that the normalization method does not introduce artifacts into the F0 patterns that are phonologically conditioned, which lies at the core of this work.

In this study, F0 scaling and F0 standardization are contrasted as these two normalization methods have both been used in previous studies. F0 scaling refers to the method adopted in Liberman (2014), where vowel-onset F0 is scaled relative to the mean F0 of (part of) the F0 trajectory (more on this method in Section 2.2.2). F0 scaling therefore belongs to a class of normalization methods termed *mean normalization* (i.e., the ratio of  $x$  to the mean of  $x$ ; Johnson, 2020). F0 standardization refers to the within-speaker  $z$ -transformation employed in Luo (2018) and Guo (2020), which is based on general-purpose statistical techniques. The differences between the scaling and standardization approaches will be discussed more in Section 2.5.2 after the results of the study are presented. Before closing this section, it is worthwhile to mention another commonly used F0 normalization method—semitone transformation—and explain why it is not used here. The semitone trans-

formation, as operationalized here, can be related to F0 in Hz via the following formula (Clayards, 2018; Dmitrieva et al., 2015; Shultz et al., 2012):

$$\text{vowel-onset F0 in semitone} = \frac{12 \ln(x/\text{individual mean vowel-onset F0})}{\ln 2}.$$

The semitone transformation is thus a kind of mean normalization, but shares with  $z$ -transformation the property that a higher F0 in the original scale in the domain of normalization (e.g., within individual speakers) will be mapped to a higher F0 in the transformed scale in the same domain. Therefore, analyses based on  $z$ -score and those based on semitones will derive the same pattern. The reason to avoid using semitones here is chiefly analytical. That is, because  $z$ -transformation squeezes most values into a range between  $-2$  and  $2$  in most situations, it is easier to place a prior on the coefficient of a predictor: the prior should result in the effect of the predictor that shifts the transformed F0 largely within the range  $[-2, 2]$ . With the semitone transformation, it is harder to predict the range of transformed F0 values, which in turn makes it harder to specify an appropriate prior for coefficients. Furthermore, Rose (2016) shows  $z$ -score normalization to give superior clustering than semitone (cf. Zhang [2018]).<sup>1</sup>

#### 2.1.4 The Current Study

The studies reviewed in Section 2.1.1 are all experimental in nature, but controlled experimental studies represent but one way to tap into onset F0 perturbation in Mandarin—another way to examine this phenomenon is through corpora. One such corpus “study” on the onset-F0 perturbation in Mandarin is a blog post by Liberman (2014), which explored the effects of consonant voicing on F0 of following vowels, based on the data from the training-set part of the Mandarin Chinese Phonetic Segmentation and Tone corpus (Yuan et al., 2014; see Section 2.2.1 for more detail about the corpus). He extracted all tonal syl-

---

<sup>1</sup>The effectiveness of normalization is evaluated through the *normalization index*, which captures the intuition that between-speaker differences should be minimized after normalization. The normalization index is the ratio of the dispersion coefficients (i.e., the proportion of the overall variance that is due to between-speaker differences) for the raw and normalized data, and quantifies how much the between-speaker variance in the raw data is reduced by normalization.

lables whose initial consonants were /p, p<sup>h</sup>, t, t<sup>h</sup>, k, k<sup>h</sup>/, measured F0 over the initial 50 ms of the rime with the Entropic Signal Processing System (ESPS; <http://languagelog.ldc.upenn.edu/myl/esps60.6.linmac.src.tgz>), which is a package of commands and programming libraries for speech signal processing, and scaled the F0 trajectory of each rime with the average F0 value over the trajectory (more on the scaling procedure in Section 2.2.2). He then aggregated these F0 trajectories, element-wise, for each combination of the four tones and the six initial consonants. Visually inspecting the mean trajectories, he observed that, in general, aspirated stops led to a higher vowel-onset F0 than unaspirated stops, which puts his results in line with Howie (1976), Chen (2011), and Luo (2018). However, the difference in F0 appeared to be contingent on tone. Tone 1 and Tone 2 showed a similar difference, but the difference seemed to be absent in Tone 3 and rather attenuated in Tone 4. As he did not conduct statistical tests on the data, it is not yet clear if the reported difference was statistically significant—this was addressed in the current study.

Another corpus study was reported in Sonderegger et al. (2017), which investigated the cross-linguistic effect of vowel height and preceding consonant voicing on F0 among 14 languages, with Mandarin being one of the languages. In the following, I review the results concerning the effect of consonant voicing in Mandarin. The Mandarin tokens came from around 20 hours of read sentences from the GlobalPhone corpus (Schultz, 2002). In particular, Sonderegger et al. (2017) used Praat (Boersma and Weenink, 2021) to extract F0 contours for all /a, i, u/ vowel tokens in utterance-initial obstruent-vowel syllables. The F0 values were then transformed to semitones (within-speaker), and the mean F0 over the initial 50 ms of a vowel was fit with a linear mixed-effects model to assess the effect of obstruent voicing. The model included fixed effects of consonant manner and voicing class, vowel height, and tone, while the random structure contained random intercepts for speaker, word, and following segment, and by-speaker random slopes for vowel height and consonant voicing. The results showed a positive trend, where aspirated obstruents were followed by a higher vowel-onset F0 than unaspirated counterparts, although large within-language inter-speaker variability was also observed. These results hence are similar to those of Liberman (2014), though Sonderegger et al. (2017) did not examine the effect of lexical tone.

Inspired by the two corpus studies just reviewed, this study also focuses on whether and how the phonological voicing of the initial consonant has an impact on vowel-onset F0 in Mandarin. In addition, it attempted to address some of the issues raised in the post, such as the lack of statistical analyses and of controlling of other factors that might also affect F0. This study therefore complements previous experimental investigations by using a corpus-based approach, and extends Liberman's (2014) original work by incorporating more rigorous statistical analyses. Compared with Liberman (2014), I included data from both the training-set and test-set parts of the corpus (see Section 2.2.1) and broadened the categories of initial consonants to include sonorants /m, n, l/, as these consonants were also considered in Luo's (2018) experimental study investigating vowel-onset F0 in Mandarin. On the other hand, I narrowed in on the onset-F0 perturbation only in Tone 1 and Tone 4 syllables for both theoretical and practical reasons. Theoretically, the tonal contours for both Tone 1 and Tone 4 start in a high F0 range (as reviewed in Section 2.1.1), so their results should be more comparable to each other. Indeed, the experimental studies reviewed above have shown that these two tones tended to pattern together in terms of their vowel-onset F0 behavior (Chen, 2011; Guo, 2020; Luo, 2018; Xu and Xu, 2003). These two tones are also the major players in the experimental studies presented in the subsequent chapters. Practically speaking, because Tone 1 and Tone 4 begin with a higher F0 register, as compared with Tone 2 and Tone 3, automatic F0-tracking algorithms tend to give more accurate F0 estimates at the beginning of the two tones (Szakay and Torgersen, 2019). We can therefore be more confident that whatever conclusion drawn from the data is less likely simply due to F0-extraction artifacts.

Another aspect this study focuses on is the impact of F0-normalization strategies. As stated in Section 2.1.3, F0 scaling and F0 standardization are contrasted here. To foreshadow the outcomes, the two F0-normalization methods did lead to different results. While the results based on F0 scaling were largely aligned with what has been stated in Liberman (2014), with aspirated stops correlated with a higher vowel-onset F0 than unaspirated stops, the results derived from F0 standardization suggested otherwise. The cause and implications of the diverging results are discussed in Section 2.5.

## 2.2 Data and Methods

### 2.2.1 Corpus Data

The data came from the Mandarin Chinese Phonetic Segmentation and Tone corpus (Yuan et al., 2014), which contains the phonetic segmentation and tone labels from 7,849 Mandarin utterances produced by 20 native speakers (7 female, 13 male; information on age was not provided; see Table C.1 in the appendix for the gender of each speaker). The utterances, defined as between-pause units in the corpus, were in turn derived from the Mandarin Broadcast News Speech and Transcripts corpus (Huang et al., 1997), which consists of approximately 30 hours of Mandarin broadcast news recordings from Voice of America, China Central TV, and KAZN-AM, a commercial radio station based in Los Angeles, CA.

Yuan et al.’s (2014) corpus consisted of two sets: a training set of 7,539 utterances and a test set of 300 utterances. The phonetic segmentation and transcription of the training set were automatically obtained using a Hidden Markov Model based forced-aligner trained on the same utterances, while those of the test set were manually annotated. The accuracy of the trained forced-aligner was checked against the test set, and an accuracy of 93.1% was reported for agreement of alignment within 20 ms, with a 1.2% rate of wrong assignment of tones.

### 2.2.2 Dataset Construction

Two datasets—F0 scaling and F0 standardization—were constructed, and the two datasets only differed in the way vowel-onset F0 was normalized, as explained below. The preprocessing steps described below applied to both datasets, unless otherwise indicated.

All Tone 1 and Tone 4 syllables beginning with any of the following consonants were extracted: /p<sup>h</sup>, p, m, t<sup>h</sup>, t, n, l, k<sup>h</sup>, k/. This yielded 17,174 items in total. For each extracted syllable, the following information was then determined or measured, as will be elaborated on immediately below: the lexical tone of the syllable (Tone 1 or Tone 4), its position within the utterance (utterance-initial or utterance-medial), the place of articulation of the initial consonant (labial, alveolar, or velar), the laryngeal quality of the initial consonant (aspirated, unaspirated,

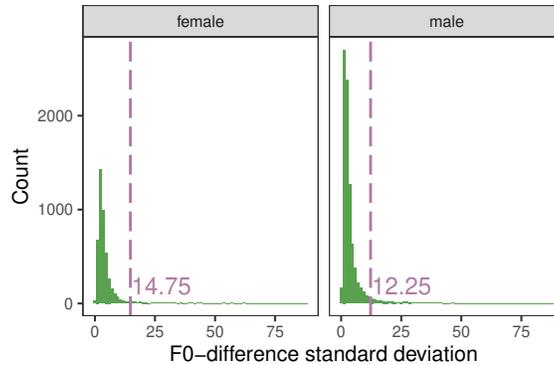
or sonorant), the height of the vowel/glide immediately following the initial consonant (high or non-high), consonant duration (i.e., closure duration, burst, and voice onset time combined for an oral stop), and vowel-onset F0.

The information on lexical tone was retrieved directly from the transcription, as lexical tone was marked for each syllable. Utterance-initial was defined as the position immediately follows a pause, and utterance-medial otherwise. Labials included /p<sup>h</sup>, p, m/, alveolars included /t<sup>h</sup>, t, n, l/, and velars included /k<sup>h</sup>, k/. In terms of laryngeal quality, aspirated consonants referred to /p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>/, unaspirated consonants to /p, t, k/, and sonorant consonants to /m, n, l/. If the initial consonant was followed by /i, y, u, j, ɥ, w/, the token was coded as high, and non-high otherwise. Consonant duration of an oral stop was operationalized as the duration of the corresponding segment, calculated from the boundaries set by the forced aligner. The measurement and quantification of vowel-onset F0 involved multiple steps, as well be explained below. It should be noted that more preprocessing steps were involved in this study than has been taken in Liberman (2014), where F0 scaling was applied to raw F0 estimates without the intermediate processing steps.

1. **F0 estimation:** The Robust Epoch And Pitch Estimator (REAPER, <https://github.com/google/REAPER>) was used to estimate the F0 trajectory over the entire utterance. The F0 search floor was set to 20 Hz (the default is 40 Hz), and the F0 search ceiling was defaulted at 500 Hz; the other parameters were kept at default values. The decision to expand the F0 search range was to ensure that the F0 of the low target of Tone 4 could be estimated more accurately, though this particular setting was not likely to affect the results, as only the first 50 ms of the vocalic part was analyzed (see below). The output file had three columns. The first column listed the timestamps at which F0 values were estimated, with the interval between consecutive timestamps being 1 ms. The second column indicated whether voicing was detected for a particular timestamp (1 for yes, 0 for no). The third column detailed the estimated F0 value in hertz (Hz) for each timestamp (−1 was used for the timestamps where no voicing was detected). All the subsequent preprocessing steps before F0 normalization were done based on F0 measures in Hz.
2. **F0 extraction:** Because of the F0 normalization strategy adopted by Liber-

man (2014) (explained below), extracted Tone 1 and Tone 4 syllables shorter than 50 ms were discarded ( $n = 365$ ). For each retained syllable, 10 F0 values were taken at 10 equidistant points, each 5 ms apart, starting from the vowel onset. Because the timestamps at which REAPER estimated F0 did not necessarily coincide with the 10 equidistant points (e.g., the vowel onset was at 0.8225 s, but REAPER estimated F0 at 0.822 s and 0.823 s), the F0 value at each of the 10 equidistant points was obtained through interpolation using the F0 values estimated by REAPER at the two nearest timestamps. If the estimated F0 value by REAPER was  $-1$  at any timestamps used for interpolation, the corresponding token was also excluded ( $n = 3,620$ ).

3. **Oddball removal:** The procedure above resulted in a sequence of 10 F0 estimates for each vowel. To further improve the quality of the dataset, tokens with irregular F0 fluctuations, which were assumed to be F0 tracking errors, were removed. Irregular F0 fluctuations were characterized by the standard deviation of F0 changes, as explained below. For each token, a standard deviation was computed for differences between adjacent F0 values. For instance, a token with the following F0 estimates— $[174, 173, 169, 163, 154, 148, 148, 145, 143, 134]$ —would have a sequence of 9 F0 differences  $[174 - 173, 173 - 169, 169 - 163, 163 - 154, 154 - 148, 148 - 148, 148 - 145, 145 - 143, 143 - 134] = [1, 4, 6, 9, 6, 0, 3, 2, 9]$  and an F0-difference standard deviation of 3.28. Using this metric, tokens with huge F0 fluctuations would have greater standard deviations. Next, the 95% quantile of these standard deviations was identified, separately for male and female speakers. The by-gender distributions of these standard deviations and the 95% quantiles are plotted in Figure 2.1. Tokens whose F0-difference standard deviation exceeded the gendered 95% quantile were then discarded ( $n = 660$ ).
4. **F0 normalization:** This step differed, depending on the dataset.
  - **For the F0-scaling dataset:** This normalization strategy followed that documented in Liberman (2014), which represents a larger class of normalization methods that compare values to a reference level (e.g., the



**Figure 2.1:** Distributions of F0-difference standard deviation, separated for female and male speakers. The vertical dashed lines mark the 95% quantile for each gender group.

$\Delta F$  method [Johnson, 2020] for vowel formant normalization). F0 scaling therefore roughly quantifies how quickly an F0 contour descends from the vowel onset. For each retained token, the 10 sequential F0 estimates were then divided by their average value. For example, the 10 sequential F0 estimates might be [174, 173, 169, 163, 154, 148, 148, 145, 143, 134]. The mean value is 155.1, and so the scaled values are [1.12, 1.12, 1.09, 1.05, .99, .95, .95, .93, .92, .86]. Therefore, a value of 1.12 means that the value at that point was 12% greater than the average of the values over the first 50 ms. Finally, only the scaled F0 values at the first point (e.g., 1.12 in the above sequence) were kept in the dataset as the current study focuses on the vowel-onset F0; the temporal aspect of F0 perturbation is left for future research.

- **For the F0-standardization dataset:** Only the F0 estimate at the first point was involved in this step; the other nine F0 estimates were discarded. All the first F0 estimates were grouped by speaker and then standardized within each speaker:  $[F0 - \text{Mean}(F0)] / \text{SD}(F0)$ .

Each kept token was also assigned a word index to be used in the statistical analyses (see Section 2.2.3). A word index was created for each unique combination of an initial, a final, and a tone. This approach meant that homophones were

conflated into a single word type. For example, the homophones *ti1* 踢 ‘to kick’ and *ti1* 梯 ‘ladder’ were treated as a single word type and assigned the same word index. This approach was taken because the forced-aligned results are given in Pinyin, instead of Chinese characters, in the corpus.

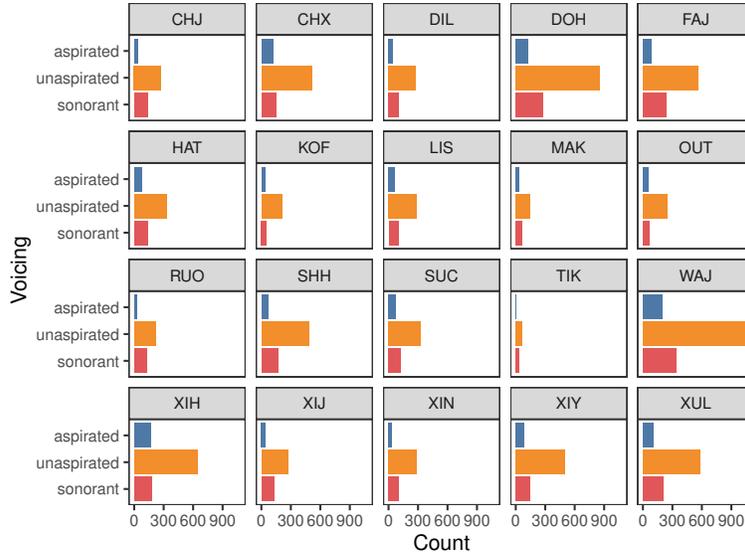
Altogether, the processing procedure resulted in a dataset of 12,529 tokens from 201 word types. The number of tokens each speaker contributed, broken down by voicing, is plotted in Figure 2.2. Figure 2.3 shows the distributions of scaled vowel-onset F0 by utterance position, lexical tone, vowel height, and stop category, and Figure 2.4 shows the distributions of standardized vowel-onset F0 by the same factors. Summary statistics for consonant duration and vowel-onset F0 are given in Table C.2 in the appendix.

Comparing Figure 2.3 and Figure 2.4, the two ways of normalization seem to show somewhat different results. For example, focusing on the vowel-onset F0 difference between aspirated and unaspirated stops, there are more pairs where the median F0 of aspirated stops is higher than that of unaspirated stops in the case of F0 scaling. For F0 standardization, the number of pairs where aspirated stops leading to a higher median F0 than unaspirated stops is approximately the same as those where the opposite pattern holds. In addition, neither methods of normalization shows a consistent aspirated > unaspirated > sonorant or unaspirated > aspirated > sonorant pattern across the board. However, these visual trends should not be over-interpreted as some groups have very few tokens.

### 2.2.3 Statistical Analyses

Two sets of analyses—one on the F0-scaling dataset and the other on the F0-standardization dataset—were performed. Since the logic behind both sets of analyses was the same, the description below applied to both cases, unless otherwise stated.

Given the current study’s focus on vowel-onset F0, I modeled vowel-onset F0 as a linear function of a number of variables that were properties of tokens and speakers. The names of predictor variables are given in **boldface**, and different levels within a variable are indicated in SMALL CAPS. In what follows, I describe the variables considered in the model comparison process, summarize the model



**Figure 2.2:** The number of tokens each speaker contributed to the dataset, broken down by phonological voicing.

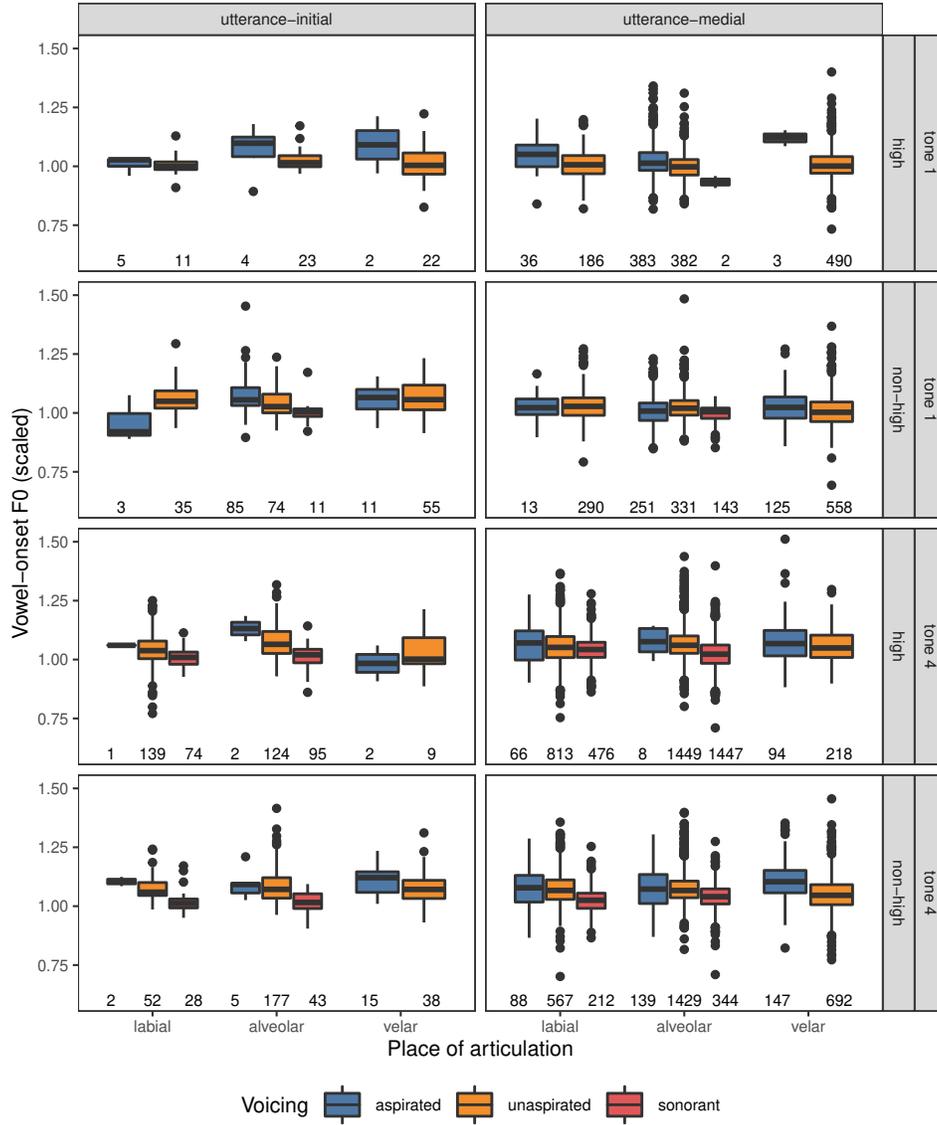
structure in terms of the priors used for parameters, explain the inference criteria, and finally move on to the model candidates considered in model comparison.

### Variables

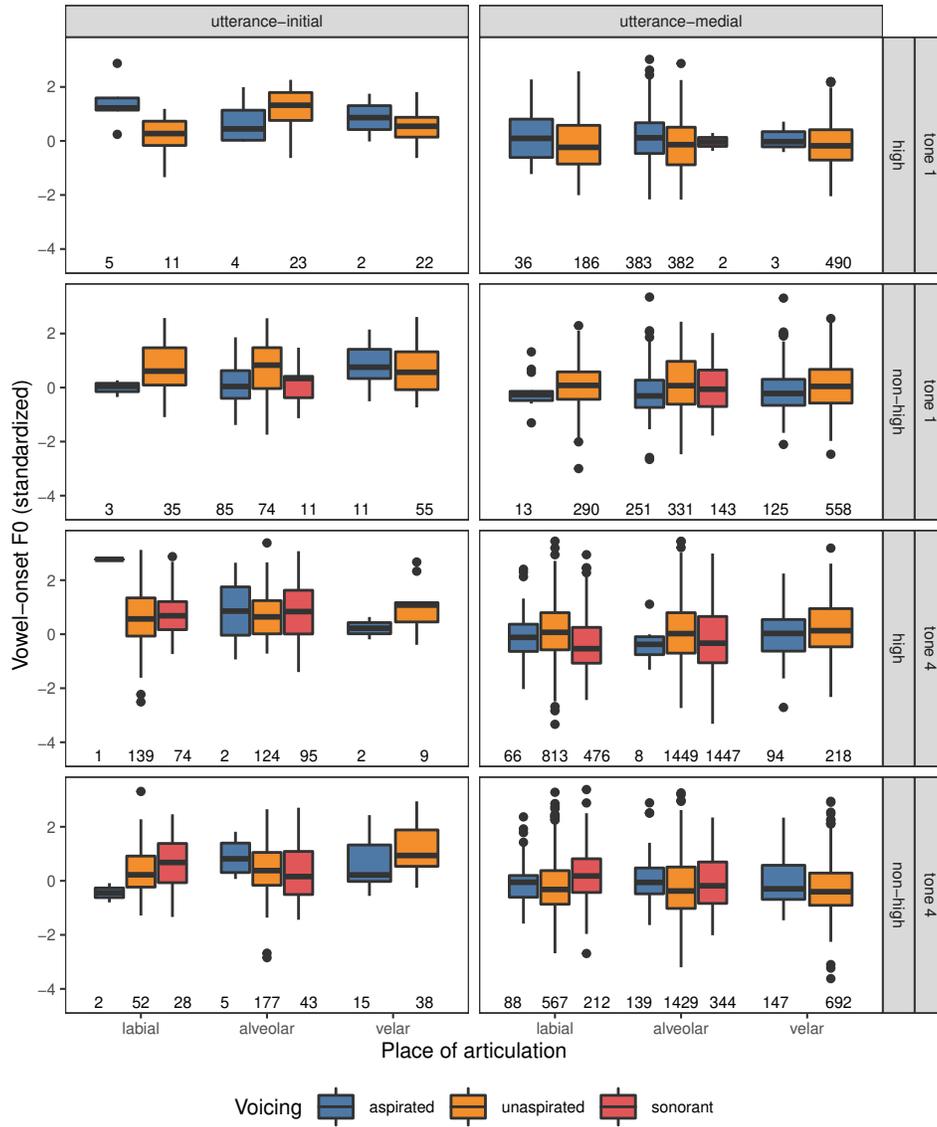
One utterance-level predictor—**position**—was included in the model. This predictor was coded using sum contrasts, with `UTTERANCE-INITIAL` coded as 1 and `UTTERANCE-MEDIAL` as  $-1$ .

Four token-level predictors were also considered in model comparison (see the Candidate Models section below): **tone**, **height**, **place of articulation (PoA)**, and **voicing**. Sum contrasts were used for **tone** (`TONE 1`, `TONE 4` =  $[1, -1]$ ), **height** (`HIGH`, `NON-HIGH` =  $[1, -1]$ ), and **PoA** (`LABIAL`, `ALVEOLAR`, `VELAR`, with `VELAR` coded as  $-1$ ).<sup>2</sup> For **voicing**, forward-difference coding was used, correspond-

<sup>2</sup>Instead of sum coding, another possible coding scheme for these categorical variables is the weighted effect coding (Nieuwenhuis et al., 2017). With weighted effect coding, the levels of a categorical variable are compared against the weighted mean. For instance, the effect for `LABIAL` of **PoA** would indicate that `LABIAL` differs from the weighted mean across all levels (i.e., `LABIAL`, `ALVEOLAR`, and `VELAR`). Weighted effect coding accounts for unbalanced observational data (e.g.,



**Figure 2.3:** Scaled vowel-onset F0 values as a function of utterance position, lexical tone, vowel height, place of articulation, and phonological voicing. The number below each “box” indicates the number of tokens in each distribution.



**Figure 2.4:** Standardized vowel-onset F0 values as a function of utterance position, lexical tone, vowel height, place of articulation, and phonological voicing. The number below each “box” indicates the number of tokens in each distribution.

ing to ASPIRATED vs. UNASPIRATED and UNASPIRATED vs. SONORANT.

Two factors—**speaker** and **word**—were included in the models as part of random effects. The **speaker** variable had 20 levels, while the **word** variable had 201 levels.

The dependent variable in all models was normalized vowel-onset F0, as discussed above.

### Model Structure

Vowel-onset F0 was modeled as a function of a subset of the predictor variables introduced above, using Bayesian linear mixed-effects models. All models were fitted using the *brm* function from the *brms* R package (Bürkner, 2017, 2018, 2021). The *brms* package provides an interface to the Stan probabilistic programming language (Stan Development Team, 2020) and adopts the formula specification of models from the *lme4* R package (Bates et al., 2015). One important reason that Bayesian models are preferred to frequentist models is that the former allow for graded statements about the strength of evidence for all parameters.

All candidate models shared general specifications. Main-effect terms were included for the predictor variables selected in a particular candidate model. As one of the goals of this case study is to examine the influence of phonological voicing on vowel-onset F0, two-way interaction terms considered all had **voicing** as a component (i.e., **voicing** × **PoA** or **voicing** × **height**). I did not consider any three-way interactions as they are in general harder to interpret and could dramatically slow

---

more utterance-medial tokens than utterance-initial tokens) and facilitates the interpretation of effects (te Grotenhuis et al., 2017b). However, interaction terms involving variables with weighted effect coding cannot be created simply by multiplying the values of the variables that make up the interaction as the coding matrix for the interaction is a function of the numbers of observations of the variables that interact (te Grotenhuis et al., 2017a). The function `wec.interact` from the *wec* R package (Nieuwenhuis et al., 2017) currently only supports the interaction between two weighted effect coded variables, and between a weighted effect coded variable and a continuous variable. It is still not clear how to form an interaction term between a forward-difference coded variable (e.g., **voicing** in the present study) and a weighted effect coded variable. This kind of interaction is needed in the model, as described in Section 2.2.3. Due to this limitation, I still resorted to sum coding for all analyses reported in this chapter. In Table C.3 and Table C.4 in the appendix, however, I included the output from the model  $F0 \sim \text{position} + \text{height} + \text{tone} + \text{voicing} + \text{PoA} + (1 + \text{position} + \text{height} + \text{tone} + \text{voicing} + \text{PoA} \parallel \text{speaker}) + (1 + \text{position} \parallel \text{word})$  with **position**, **height**, **tone**, and **PoA** being weighted effect coded. The trends revealed in these models are the same as those described in the main text.

down model sampling. All models also included by-speaker and by-word random intercepts, to account for variability in vowel-onset F0 of speakers and words beyond the effects of predictor variables in the model. All possible by-speaker and by-word random slopes were also included to account for variability among speakers and words in the effects of predictors on F0 (Barr et al., 2013). Correlations between random-effect terms were omitted to avoid divergent transitions and to prevent the model from hitting the maximum treedepth.

Regularizing priors were used for the intercept and all population-level effects. For the analyses on the F0-scaling dataset,  $\text{Normal}(\mu = 1, \sigma = 1)$  was used for the intercept, and  $\text{Normal}(\mu = 0, \sigma = 1)$  was used for population-level effects. For the analyses on the F0-standardization dataset,  $\text{Normal}(\mu = 0, \sigma = 1)$  was used for both the intercept and population-level effects. For both sets of analyses, the prior for the individual-level standard deviations was an exponential( $r = 1$ ) distribution. All models were run with four chains, with each chain having 2,000 iterations (including 1,000 warm-up iterations). All models showed no divergent transitions, and had  $\hat{R}$  values close to 1 (i.e., all  $\hat{R} < 1.01$ ), which indicates that chains were well-mixed.

### **Inference Criteria**

Evidence embedded in each model was evaluated in two ways: (i) the posterior distributions of parameters, and (ii) comparison of models of different complexities. Specifically, I consider there to be strong evidence for a non-null effect if the 89% credible interval (CrI)—the narrowest interval that contains 89% of the posterior density—for the parameter does not include 0. If the 89% CrI spans 0 but the probability of the parameter not changing direction is at least 89%, I consider this to represent weak evidence for a given effect. The decision to use CrIs of 89%, as opposed to 95%, is based on Koster and McElreath (2017) and McElreath (2020) and to discourage the association between a Bayesian posterior distribution and a  $p$ -value. However, when the posterior distributions of parameters are presented in a figure, the CrIs of 95% are also plotted. Model comparison was done by means of the Bayesian leave-one-out estimate of expected log pointwise predictive density (ELPD-LOO; Vehtari et al., 2017), which aims to gauge a model's *predictive accu-*

racy. A higher ELPD-LOO value means the model has a better predictive accuracy. The results from model comparison thus inform us whether a variable contributes substantially to a model’s predictive power. Following Sivula et al. (2020), when the estimated absolute difference in ELPD-LOO between two models is at least 4, and 0 is not within two standard errors of the estimated difference, there is evidence that the two models give different predictions.

In the following sections, model parameters are reported in terms of marginal posterior means of parameters, 89% CrIs, and the probability of effect direction.

### **Candidate Models**

The construction of candidate models relied both on prior knowledge about factors affecting vowel-onset F0 and on a compromise between model complexity and predictive accuracy. All candidate models on the F0-scaling dataset are given in Table 2.2, and those on the F0-standardization dataset are given in Table 2.3. Given that a word’s position in the utterance (“downdrift”; Connell, 2001), vowel height (“intrinsic F0”; Whalen and Levitt, 1995), and tonal categories might all affect F0, the base model (i.e., M1) included **position**, **height**, and **tone** as control predictors. As one of the goals is to establish whether and how vowel-onset F0 might be influenced by phonological voicing, further models were constructed by incrementally adding terms that involved **voicing**. For example, the comparison between M1 and M2 assessed the contribution of voicing in predictive accuracy, and comparing M3 and M6 examined the importance of the interaction between voicing and vowel height in predicting vowel-onset F0 values. Furthermore, **place of articulation** was also considered in order to examine its contribution to vowel-onset F0.

**Table 2.2:** Candidate models on scaled vowel-onset F0 considered in the model comparison, with their ELPD-LOO means and standard deviations. An intercept was included in each model but is omitted here to save space.

Model	ELPD-LOO	ELPD-LOO standard error	Predictors
M1	16203.1	117.4	position + height + tone + (1 + position + height + tone    speaker) + (1 + position    word)
M2	16353.4	117.4	position + height + tone + voicing + (1 + position + height + tone + voicing    speaker) + (1 + position    word)
M3	16438.1	117.3	position + height + tone + voicing + PoA + (1 + position + height + tone + voicing + PoA    speaker) + (1 + position    word)
M4	16441.1	117.2	position + height + tone + voicing + PoA + voicing × PoA + (1 + position + height + tone + voicing + PoA + voicing × PoA    speaker) + (1 + position    word)
M5	16440.7	117.6	position + height + tone + voicing + PoA + voicing × tone + (1 + position + height + tone + voicing + PoA + voicing × tone    speaker) + (1 + position    word)
M6 (final)	16451.0	117.2	position + height + tone + voicing + PoA + voicing × height + (1 + position + height + tone + voicing + PoA + voicing × height    speaker) + (1 + position    word)
M7	16452.9	117.3	position + height + tone + voicing + PoA + voicing × height + voicing × position + (1 + position + height + tone + voicing + PoA + voicing × height + voicing × position    speaker) + (1 + position    word)

**Table 2.3:** Candidate models on standardized vowel-onset F0 considered in the model comparison, with their ELPD-LOO means and standard deviations. An intercept was included in each model but is omitted here to save space.

Model	ELPD-LOO	ELPD-LOO standard error	Predictors
M1	-16867.8	79.0	position + height + tone + (1 + position + height + tone    speaker) + (1 + position    word)
M2	-16855.9	79.1	position + height + tone + voicing + (1 + position + height + tone + voicing    speaker) + (1 + position    word)
M3 (final)	-16846.1	79.2	position + height + tone + voicing + PoA + (1 + position + height + tone + voicing + PoA    speaker) + (1 + position    word)
M4	-16841.9	79.4	position + height + tone + voicing + PoA + voicing × PoA + (1 + position + height + tone + voicing + PoA + voicing × PoA    speaker) + (1 + position    word)
M5	-16847.1	79.3	position + height + tone + voicing + PoA + voicing × tone + (1 + position + height + tone + voicing + PoA + voicing × tone    speaker) + (1 + position    word)
M6	-16838.6	79.3	position + height + tone + voicing + PoA + voicing × height + (1 + position + height + tone + voicing + PoA + voicing × height    speaker) + (1 + position    word)
M7	-16848.2	79.2	position + height + tone + voicing + PoA + voicing × position + (1 + position + height + tone + voicing + PoA + voicing × position    speaker) + (1 + position    word)

## 2.3 Results: Scaled F0

The ELPD-LOO mean and standard error for each candidate model are listed in Table 2.2. As stated above, a higher ELPD-LOO value means the model has a better predictive accuracy, so, for instance, M2 makes better predictions than M1. Finally, model comparison results are summarized in Table 2.4, in terms of difference in ELPD-LOO values and associated standard errors. Note that the difference score in each cell was computed by subtracting the ELPD-LOO value of the model represented in the column from the ELPD-LOO value of the model indicated in the row. For instance, the difference  $-150.2$  came from  $ELPD-LOO_{M1} - ELPD-LOO_{M2} = 16203.1 - 16353.4$ .

Model comparison confirmed the importance of phonological voicing in conditioning vowel-onset F0 (i.e., M1 vs. M2) and spoke to the importance of interaction between voicing and vowel height (i.e., M3 vs. M6). Place of articulation also seemed to contribute to the predictive power of a model (i.e., M2 vs. M3). As no significant gain in predictive power was observed past M6, M6 was selected as the final model, on which the following discussion is based.

In presenting results, I summarize the output from the final model in terms of posterior distributions for key parameters. I first interpret population-level parameter estimates before moving on to individual-level estimates.

**Table 2.4:** Model comparison results for key scaled-F0 model pairs, expressed in differences in ELPD-LOO and associated standard errors. Pairs that are judged to differ in predictive power are marked by asterisks.

Model	M2	M3	M4	M5	M6	M7
M1	-150.2* (18.2)					
M2		-84.8* (14.0)				
M3			-3.0 (4.4)	-2.6 (4.5)	-12.9* (5.5)	
M6						-1.9 (3.3)

### 2.3.1 Scaled F0 at the Population Level

The population-level results of M6 are summarized in Table 2.5 and depicted in Figure 2.5. For the control predictors (i.e., **position**, **height**, and **tone**), there was evidence that high vowels and Tone 1 were associated with a lower vowel-onset F0 (height:  $\bar{\beta} = -.0031$ , 89% CrI =  $[-.0066, .0004]$ ,  $p(\beta < 0) = .92$ ; tone:  $\bar{\beta} = -.0211$ , 89% CrI =  $[-.0251, -.0173]$ ,  $p(\beta < 0) = 1.00$ ), but there was little evidence that a word's position in the utterance had an effect ( $\bar{\beta} = .0019$ , 89% CrI =  $[-.0023, .0061]$ ,  $p(\beta > 0) = .78$ ). As for place of articulation, even though this variable improved the overall model predictive ability, it did not seem to be correlated with a systematic rise or fall in vowel-onset F0 (labial:  $\bar{\beta} = .0000$ , 89% CrI =  $[-.0042, .0041]$ ,  $p(\beta > 0) = .50$ ; alveolar:  $\bar{\beta} = -.0003$ , 89% CrI =  $[-.0056, .0048]$ ,  $p(\beta < 0) = .54$ ). Crucially for this case study, there was evidence that the phonological voicing of the initial consonant patterned with vowel-onset F0: the F0 following an aspirated stop tended to be higher than that following an unaspirated stop ( $\bar{\beta} = .0123$ , 89% CrI =  $[.0005, .0242]$ ,  $p(\beta > 0) = .95$ ), and the F0 after an unaspirated stop was in turn higher than that after a sonorant ( $\bar{\beta} = .0407$ , 89% CrI =  $[.0310, .0505]$ ,  $p(\beta > 0) = 1.00$ ). There was also evidence that high vowels supported a larger F0 difference between an aspirated and an unaspirated consonant ( $\bar{\beta} = .0101$ , 89% CrI =  $[.0024, .0177]$ ,  $p(\beta > 0) = .98$ ), although there was much weaker evidence that the F0 difference between an unaspirated and a sonorant consonant was *reduced* for a high vowel ( $\bar{\beta} = -.0053$ , 89% CrI =  $[-.0124, .0019]$ ,  $p(\beta < 0) = .88$ ).

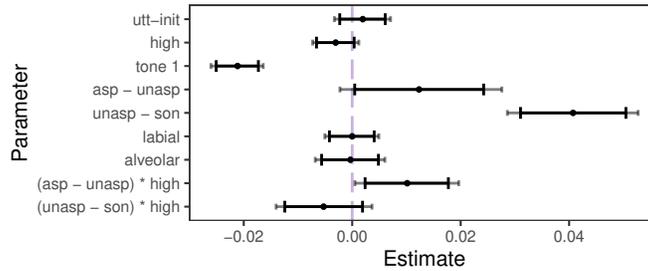
### 2.3.2 Scaled F0 at the Individual Level

The numerical values for all parameters at the individual level can be found in Table C.5 and Table C.6 in the appendix. Since this study focuses on the patterning between phonological voicing and vowel-onset F0, the following discussion is restricted to the four parameters involving the predictor **voicing**—the difference between aspirated and unaspirated stops, the difference between unaspirated stops and sonorants, and the interactions between these two parameters and vowel height. The individual-level estimates for these four parameters are depicted in Figure 2.6.

The first panel in Figure 2.6 indicates that half of the 20 speakers showed

**Table 2.5:** Marginal posterior summaries for population-level parameters from M6 (scaled F0). The contrast coding scheme for each variable is described in Section 2.2.3. The parameters whose effects are judged to be strong are marked with \*\*, and those whose effects are judged to be weak are marked with \*.

Parameter	Mean	SD	89% CrI	$p(\text{dir.})$
intercept**	1.0380	.0046	[1.0309, 1.0451]	$p(\beta > 0) = 1.00$
utt-init – (utt-init + utt-med)/2	.0019	.0026	[–.0023, .0061]	$p(\beta > 0) = .78$
high – (high + non-high)/2*	–.0031	.0022	[–.0066, .0004]	$p(\beta < 0) = .92$
tone 1 – (tone 1 + tone 4)/2**	–.0211	.0024	[–.0251, –.0173]	$p(\beta < 0) = 1.00$
asp – unasp**	.0123	.0075	[.0005, .0242]	$p(\beta > 0) = .95$
unasp – son**	.0407	.0061	[.0310, .0505]	$p(\beta > 0) = 1.00$
labial – (labial + alveolar + velar)/3	.0000	.0026	[–.0042, .0041]	$p(\beta > 0) = .50$
alveolar – (labial + alveolar + velar)/3	–.0003	.0033	[–.0056, .0048]	$p(\beta < 0) = .54$
[asp – unasp] × [high – (high + non-high)/2]**	.0101	.0048	[.0024, .0177]	$p(\beta > 0) = .98$
[unasp – son] × [high – (high + non-high)/2]	–.0053	.0045	[–.0124, .0019]	$p(\beta < 0) = .88$



**Figure 2.5:** Population-level parameters from M6 (scaled F0). Inner error bars represent 89% CrIs, and outer error bars represent 95% CrIs. The posterior mean for each parameter is indicated by a dot. Note that the intercept from M6 is omitted here as it takes relatively large values, which compress the  $x$ -axis to make the distributions of the other parameters hard to inspect.

strong evidence of their vowel-onset F0 being higher after an aspirated stop than an unaspirated one. However, four speakers also showed strong evidence for the opposite direction, with an unaspirated stop correlated with a higher vowel-onset F0 than an aspirated stop. On the other hand, all speakers show strong evidence in producing a higher vowel-onset F0 following an unaspirated stop than following a sonorant, as can be seen in the second panel. The third panel shows that some speakers consistently produced a bigger post-stop F0 difference in the context of high vowels than in the context of non-high vowels. However, as far as the posterior means are concerned, there is clear evidence that almost all speakers showed an enlarged post-stop F0 difference for high vowels. Finally, the last panel indicates an even less consistent patterns in terms of the F0 difference between an unaspirated stop and a sonorant for a high vowel. While most speakers agreed with the population pattern in having a negative mean estimate, only some speakers showed strong evidence for a reduced onset-vowel F0 difference between the two consonant types.

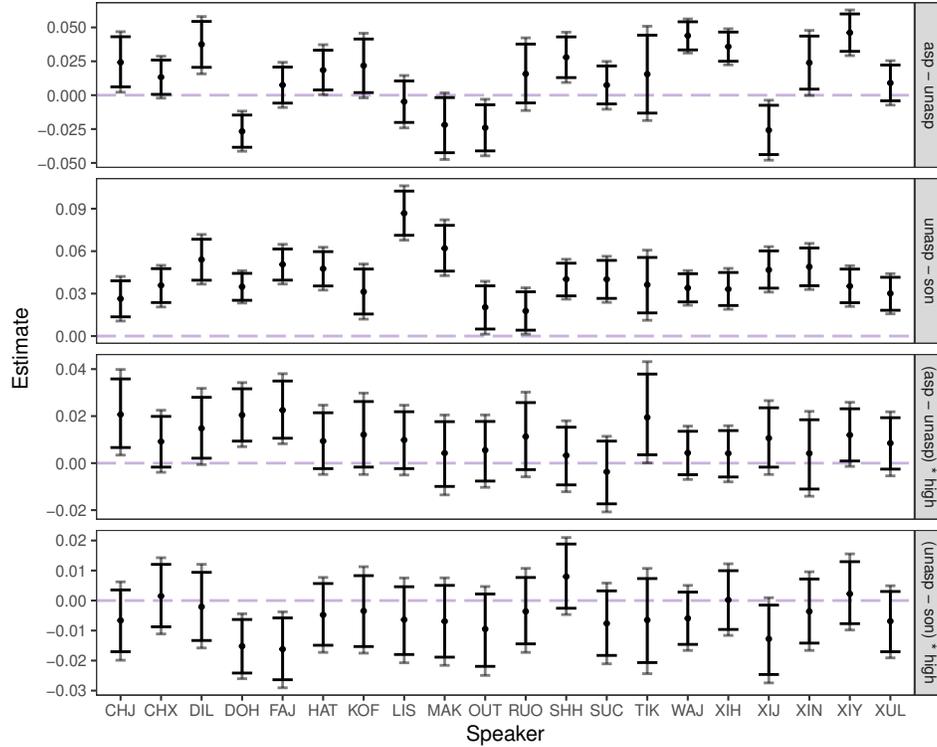
## 2.4 Results: Standardized F0

The ELPD-LOO mean and standard error for each candidate model are given in Table 2.3. Finally, model comparison results are tabulated in Table 2.6.

Similar to the results derived from the scaled-F0 dataset, model comparison attested the importance of phonological voicing (i.e., M1 vs. M2) and place of articulation (i.e., M2 vs. M3). Since no significant improvement in prediction was observed for models more complex than M3, M3 was chosen as the final model. The discussion below is therefore all based on M3.

### 2.4.1 Standardized F0 at the Population Level

Table 2.7 and Figure 2.7 summarize the population-level parameters from M3. There was strong evidence that utterance-initial position and high vowels were associated with a higher vowel-onset F0 (position:  $\bar{\beta} = .35$ , 89% CrI = [.31, .40],  $p(\beta > 0) = 1.00$ ; height:  $\bar{\beta} = .06$ , 89% CrI = [.01, .12],  $p(\beta > 0) = .96$ ). However, tone did not seem to bear an effect ( $\bar{\beta} = -.02$ , 89% CrI = [-.07, .04],  $p(\beta < 0) = .67$ ). Notice also that the direction of these effects is different from that in



**Figure 2.6:** Individual-level parameters involving the **voicing** predictor from M6 (scaled F0). Inner error bars represent 89% CrIs, and outer error bars represent 95% CrIs. The posterior mean for each parameter is indicated by a dot.

**Table 2.6:** Model comparison results for key standardized-F0 model pairs, expressed in differences in ELPD-LOO and associated standard errors. Pairs that are judged to differ in predictive power are marked by asterisks.

Model	M2	M3	M4	M5	M6	M7
<b>M1</b>	-11.9* (5.0)					
<b>M2</b>		-9.8* (4.5)				
<b>M3</b>			-4.2 (4.0)	1.0 (1.1)	-7.5 (4.7)	2.1 (1.2)

the scaled-F0 model, which suggests that Tone 1 and high vowels were associated with a lower vowel-onset F0, and that position did not have an effect. Similar to the scaled-F0 results, place of articulation also did not seem to have a systematic effect (labial:  $\bar{\beta} = -.02$ , 89% CrI =  $[-.10, .07]$ ,  $p(\beta < 0) = .64$ ; alveolar:  $\bar{\beta} = -.04$ , 89% CrI =  $[-.11, .03]$ ,  $p(\beta < 0) = .80$ ), despite its contribution to the overall predictive accuracy. Critically, and in contrast to the results on the scaled-F0 dataset, which only differs from the standardized-F0 data here in the way vowel-onset F0 was normalized, there was weak evidence that the F0 following an aspirated stop tended to be *lower* than that following an unaspirated stop ( $\bar{\beta} = -.10$ , 89% CrI =  $[-.23, .03]$ ,  $p(\beta < 0) = .90$ ). The F0 following an unaspirated stop, however, was evidently higher than that following a sonorant ( $\bar{\beta} = .19$ , 89% CrI =  $[.04, .33]$ ,  $p(\beta > 0) = .98$ ), which is similar to the scaled-F0 results.

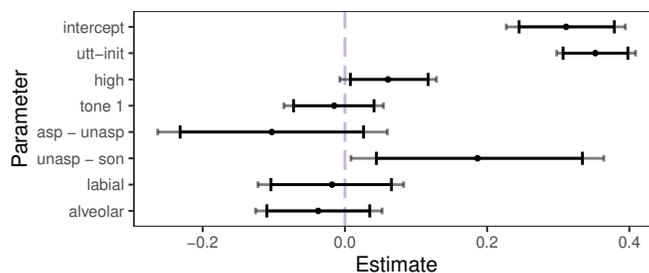
**Table 2.7:** Marginal posterior summaries for population-level parameters from M3 (standardized F0). The contrast coding scheme for each variable is described in Section 2.2.3. The parameters whose effects are judged to be strong are marked with \*\*, and those whose effects are judged to be weak are marked with \*.

Parameter	Mean	SD	89% CrI	$p(\text{dir.})$
intercept**	.31	.04	[.24, .38]	$p(\beta > 0) = 1.00$
utt-init – (utt-init + utt-med)/2**	.35	.03	[.31, .40]	$p(\beta > 0) = 1.00$
high – (high + non-high)/2**	.06	.03	[.01, .12]	$p(\beta > 0) = .96$
tone 1 – (tone 1 + tone 4)/2	–.02	.04	[–.07, .04]	$p(\beta < 0) = .67$
asp – unasp*	–.10	.08	[–.23, .03]	$p(\beta < 0) = .90$
unasp – son**	.19	.09	[.04, .33]	$p(\beta > 0) = .98$
labial – (labial + alveolar + velar)/3	–.02	.05	[–.10, .07]	$p(\beta < 0) = .64$
alveolar – (labial + alveolar + velar)/3	–.04	.05	[–.11, .03]	$p(\beta < 0) = .80$

## 2.4.2 Standardized F0 at the Individual Level

Table C.7 in the appendix lists the values for all parameters at the individual level. The following discussion, however, focuses only on the parameters involving **voicing**, the individual-level estimates of which are depicted in Figure 2.8.

The first panel in Figure 2.6 shows that five speakers evidently produced a



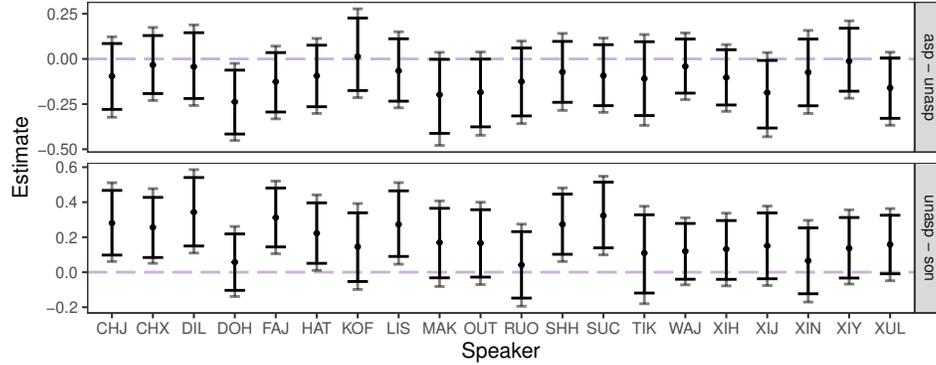
**Figure 2.7:** Population-level parameters from M3 (standardized F0). Inner error bars represent 89% CrIs, and outer error bars represent 95% CrIs. The posterior mean for each parameter is indicated by a dot.

lower vowel-onset F0 accompanying aspirated stops than unaspirated ones. For the other speakers, even though the model was not as confident, their posterior means were either around or below zero, consistent with the population-level trend that a lower vowel-onset F0 was observed following aspirated stops. The second panel contrasts the vowel-onset F0 difference between unaspirated stops and sonorants for individual speakers. The panel indicates that eight speakers showed a clear pattern of producing a higher vowel-onset F0 following an unaspirated stop. Although there was weaker evidence for the other speakers, the sign of their posterior means was aligned with the overall tendency that speakers had a higher F0 after unaspirated stops than after sonorants.

## 2.5 Discussion

### 2.5.1 Summary of Results

The current study examined the influence of phonological voicing— aspirated, unaspirated, sonorant—of the initial consonant on vowel-onset F0 in Mandarin, using the speech data from the Mandarin Chinese Phonetic Segmentation and Tone corpus (Yuan et al., 2014). Statistical analyses revealed that the patterns uncovered from the data were sensitive to how the dependent variable—that is, vowel-onset F0—was normalized. These differences are summarized in Table 2.8 and delved into below.



**Figure 2.8:** Individual-level parameters involving the **voicing** predictor from M3 (standardized F0). Inner error bars represent 89% CrIs, and outer error bars represent 95% CrIs. The posterior mean for each parameter is indicated by a dot.

**Table 2.8:** Comparison of outputs from models on the scaled-F0 and standardized-F0 datasets. The symbol  $\approx$  indicates that there is little evidence for the vowel-onset F0 to differ between the levels on the two sides of the symbol. The symbols  $<$  and  $>$  mean that there is strong evidence that vowel-onset F0 is different across the levels in the specified direction. The symbol  $\lesssim$  stands for weak evidence that the vowel-onset F0 associated with the level to the left of the symbol is smaller than that associated with the level on the right.

	Scaled-F0 model	Standardized-F0 model
<b>position</b>	UTTERANCE-INITIAL $\approx$ mean position	UTTERANCE-INITIAL $>$ mean position
<b>vowel height</b>	HIGH $\lesssim$ mean height	HIGH $>$ mean height
<b>tone</b>	TONE 1 $<$ mean tone	TONE 1 $\approx$ mean tone
<b>place of articulation</b>	LABIAL $\approx$ mean PoA	LABIAL $\approx$ mean PoA
	ALVEOLAR $\approx$ mean PoA	ALVEOLAR $\approx$ mean PoA
<b>voicing</b>	ASPIRATED $>$ UNASPIRATED	ASPIRATED $\lesssim$ UNASPIRATED
	UNASPIRATED $>$ SONORANT	UNASPIRATED $>$ SONORANT

With Liberman’s (2014) scaling method, the model suggested that aspirated consonants were correlated with a higher vowel-onset F0, as compared with unaspirated stops. This agreed with Liberman’s (2014) original observation though it should be kept in mind that a handful of speakers deviated from this trend at large. The model also suggested that unaspirated stops were linked to a higher vowel-onset F0 than sonorants, and this trend was strong at both population and individual levels.

With within-speaker F0 standardization or z-transformation, which was used in Luo (2018) and Guo (2020), a rather different picture was revealed. The model now indicated that aspirated stops were associated with a *lower* vowel-onset F0, in comparison with unaspirated stops, though the direction of F0 difference between unaspirated stops and sonorants remained unchanged. At the individual level, though there was variation, all speakers seemed to be consistent with the population-level trends, especially when posterior means were concerned.

Diverging results were also observable in the effects of other control variables. Specifically, whereas position within the utterance did not seem to be relevant for the scaled-F0 model, utterance-initial position was strongly related to a higher vowel-onset F0 in the standardized-F0 model. The effect of vowel height was also not constant among the two models: high vowels were correlated with a higher vowel-onset F0 in the standardized-F0 model but with a lower F0 in the scaled-F0 model. Finally, Tone 1 was correlated with a lower F0 in the scaled-F0 model, but not in the standardized-F0 model.

### 2.5.2 Reconciling the Differences

Since the vowel-onset F0 normalization strategy is the major difference between the two models, in this section I attempt to reconcile these differences.<sup>3</sup> I first discuss the impact of normalization on the control variables before switching gears to the properties of the dataset reflected in different normalization methods and the

---

<sup>3</sup>Of course the two models also differed in the predictor variables included, that is, the scaled-F0 model had the interaction term **voicing** × **height**, but no interaction terms were in the standardized-F0 model. However, the scaled-F0 model without interaction terms (i.e., M3 in Table 2.2) gave essentially the same output as the final scaled-F0 model, modulo a lack of estimates for interactions. The standardized-F0 model with a **voicing** × **height** interaction (i.e., M6 in Table 2.3) also led to the same conclusion as the final standardized-F0 model.

impact of normalization on the results regarding phonological voicing.

F0 scaling, as operationalized here, in fact reduces the influence of a host of factors, in addition to the gender of a speaker. That is, because the reference point for scaling is the average F0 of the trajectory in question, extrinsic influences on the absolute values of F0 (e.g., the position within the utterance) were effectively removed. Consider, for instance, a syllable in the utterance-initial position and the same syllable in an utterance-medial position. Barring the effect of focus on intonation, the utterance-initial position tends to induce a higher *absolute* F0 on the syllable than does an utterance-medial position (Connell, 2001). However, because vowel-onset F0 scaling only concerns individual F0 trajectories—that is, there is no reference to the trajectory of another syllable—the two syllables with different *absolute* F0 values might end up with similar scaled F0 values. In other words, even though the utterance position does have an impact on F0, scaling has already partialled out at least some of its impact.

Within-speaker F0 standardization/z-transformation, however, does not have quite the same effect. While it is true that standardization shifts all means to 0 and adjusts all standard deviations to 1—and thereby partials out some speaker-specific idiosyncrasies—this normalization method does not alter the relative relationship between two tokens. Using the utterance-position examples from above for illustration, the vowel-onset F0 value from the syllable with an overall higher absolute F0 will still have a higher standardized vowel-onset F0 than the syllable with a lower absolute F0, after standardization. The fact that standardization refers to the F0 values from all the other tokens for normalization means that standardization can shift individual F0 values but can never flip their relative positions. In other words, the effect of utterance position will persist even after standardization.

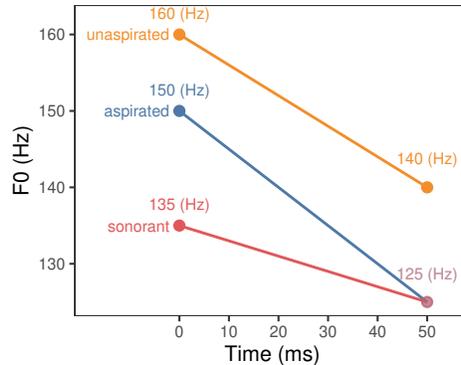
This normalization account can explain why **position** only had an effect in the standardized-F0 model (i.e., utterance-initial position was correlated with a higher F0, which was expected) but not in the scaled-F0 model. That is, F0 scaling had already removed the effect associated with **position**, so it would not show any effect. The normalization account can also partially clarify why vowel **height** had a strong effect in the standardized-F0 model but only a weak effect in the scaled-F0 model: vowel height is expected to condition *absolute* F0. However, it is not clear why it had a negative effect, as opposed to a lack of effect, in the scaled-

F0 model—more nuanced analyses are required to resolve this point. As to the differing effects of **tone**, the normalization account can also elucidate the cause, with further consideration of the influence of tonal contour on scaling. Recall from the discussion in Section 2.1.1 that Tone 1 is high-level, and Tone 4 is high-falling. The distinct contours associated with the two tones have practical implications for F0 scaling: all else being equal, we would expect Tone 1 tokens to have a smaller scaled F0 values than Tone 4 on average. Consider a case where a Tone 1 syllable and a Tone 4 syllable begin with an identical F0 value at vowel onset. For Tone 1, being a level tone, it is likely that the syllable sustains similar F0 values to the starting F0 value. This sustained, relatively flat F0 contour means that the average F0 over the initial part of the syllable will take a value close to that of the vowel-onset F0, and, as a result, the scaled F0 for Tone 1 should also hover around 1. For Tone 4, which is high-falling, however, F0 is likely to start dropping from vowel onset, so the average F0, even over the initial portion of the vowel, will assume a value smaller than that at vowel onset. As a consequence of this lower average F0 is a *higher* scaled vowel-onset F0, compared with the scaled F0 from a Tone 1 syllable sharing the same absolute vowel-onset F0. This interaction between scaling and tonal contour therefore provide an account for Tone 1 being lower in F0 (i.e., having negative parameter estimates) in the scaled-F0 model. As for Tone 1's behavior in the standardized-F0 model, we would expected Tone 1 to have a lower (standardized) F0, given that phonetically Tone 1 is typically initiated with a slightly lower F0 than Tone 4 in the citation form (Xu, 1997). The fact that **tone** lacked an effect in the standardized-F0 model might be due to the tokens being extracted from connected speech, so F0 was subject to influences from prosody, tonal coarticulation, focus, among others. It is possible that in connected speech, Tone 1 and Tone 4 have a similar onset F0, so **tone** will not have an effect. Alternatively, Tone 1 could still have a phonetically lower onset F0 than Tone 4 in connected speech, but the other predictor variables in the current model did not adequately capture variation (e.g., there was no predictor pertaining to focus in the model, but focus is known to condition F0 [e.g., Zellers and Post, 2009]), so the effect of **tone** was swamped by the variation due to other uncontrolled factors. Again, more refined analyses are required to clarify this point.

Perhaps the most unexpected contrast between the two models is the revers-

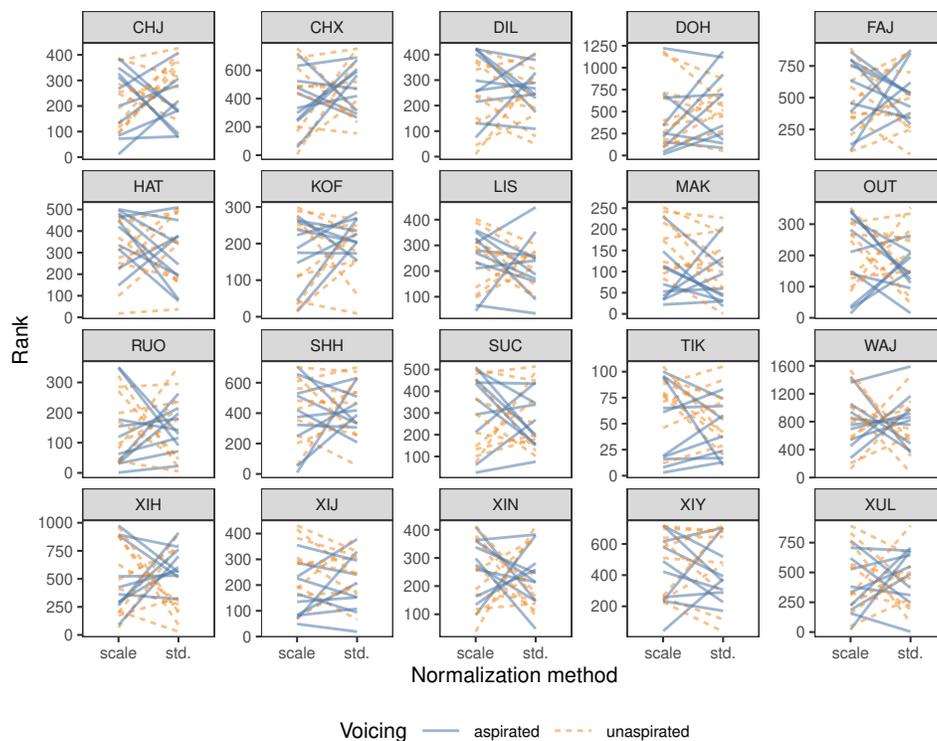
ing of the direction for the effect of aspiration. Again, the normalization account can potentially pave the path to resolve the conflicting outcomes. This conflicting situation and a possible cause are illustrated in Figure 2.9. In the particular configuration shown in the figure, standardization of vowel-onset F0 will lead to the conclusion that UNASPIRATED > ASPIRATED > SONORANT (because standardization does not change the relative relationship between elements), whereas scaling of vowel-onset F0 to average F0 will lead to the conclusion that ASPIRATED (1.09) > UNASPIRATED (1.07) > SONORANT (1.04). The extent of this pattern-flipping with the current dataset can be glimpsed by comparing the ranks of individual tokens across the two normalization methods. Scaling and standardization result in normalized F0 in different scales, rendering the two methods not directly comparable. To facilitate the cross-scale comparison, I first ranked individual tokens in terms of their normalized vowel-onset F0 within each speaker for each normalization frame, so the same token for a particular speaker would have two ranks—one in the F0-scaling approach and the other in the F0-standardization approach. I then randomly selected 10 aspirated and 10 unaspirated tokens for each speaker and compared each token’s two ranks across the two normalization methods. This comparison is depicted in Figure 2.10. The figure shows that, for every speaker, there were tokens whose ranks changed drastically across the two normalization strategies—some tokens that were ranked low in scaling were ranked high in standardization, and vice versa. Given these dramatic shifts in ranking, it is perhaps less surprising that different conclusions were reached using the datasets that only differed in the F0-normalization strategy.

In fact, upon inspecting Figure 2.9, it should become clear that the two normalization methods capture different “realities” of the same dataset. In particular, F0 scaling roughly quantifies how steeply an F0 contour falls from the vowel onset to the rest of the syllable. In this view, the finding that, with F0 scaling, aspirated tokens have a higher vowel-onset F0 than unaspirated tokens essentially reveals that aspirated tokens tend to start with a higher F0 relative to the rest of their contour than unaspirated tokens. F0 standardization, on the other hand, describes a token’s vowel-onset F0 in comparison to all the other tokens. Therefore, the finding that, with F0 standardization, aspirated tokens are associated with a lower F0 than unaspirated tokens suggests that aspirated tokens tend to have a vowel-



**Figure 2.9:** Illustration of a case where F0 standardization leads to UNASPIRATED > ASPIRATED > SONORANT, but F0 scaling leads to ASPIRATED (1.09) > UNASPIRATED (1.07) > SONORANT (1.04).

onset F0 that is lower than the average vowel-onset F0 of unaspirated tokens. The F0-standardization method then seems to be more appropriate for addressing the raised research questions. In fact, situating the current study in a broader body of works on post-stop F0 perturbation in tonal languages (e.g., Zee [1980] and Francis et al. [2006] for Cantonese, Chen [2011] for Shanghainese, Lai et al. [2009] for Taiwanese, Gandour [1974] for Thai, Phuong [1981] for Vietnamese, and Kirby and Ladd [2016] for Central Thai and North Vietnamese), F0 standardization, or other normalization strategies that do not change the relative positions among tokens, such as the cents conversion (Francis et al., 2006) or semitone transformation (Kirby and Ladd, 2016), appears to be the norm. Following this methodological tradition, the proper conclusion should then be that aspirated stops are in general associated with a lower vowel-onset F0 than unaspirated stops, at least for the speech from the selected corpus. This conclusion, however, does not mean that F0 scaling is not a useful normalization method. For example, comparison of the two measures might be important when we attempt to understand whether it is the cross-token F0 comparison or the within-token F0 contour that matters in the case of tonogenesis. The two measures also provide a way to model how listeners might perceive the initial pitch of a token: for example, will a falling F0 contour be perceived as starting with a lower F0 than a level contour held at the same F0 value as the initial F0 of the falling contour? In sum, the two normalization methods in



**Figure 2.10:** Comparison of normalized F0 values in terms of within-speaker ranks across the scaling and standardization strategies. For each speaker, 10 aspirated and 10 unaspirated tokens were randomly sampled and shown here. The two ranks belonging to the same tokens are connected by a line segment.

question profile distinct aspects of an F0 contour, and which method is appropriate depends on the research questions raised.

## 2.6 Conclusion

This study revisited the debate on the vowel-onset F0 perturbation in Mandarin, with data from a corpus of Mandarin broadcast news speech, thereby complementing previous experimental investigations on the same issue. Two datasets differing only in how vowel-onset F0 was normalized were constructed from the corpus. In one dataset, following Liberman (2014), each token’s vowel-onset F0 was scaled

relative to the mean F0 of the initial 50 ms of the vowel in the same token, and so the scaled F0 took a value in  $(0, \infty)$ . In the other dataset, following the approach adopted by a number of studies on similar research questions (e.g., Chen, 2011; Guo, 2020; Luo, 2018), each token's vowel-onset F0 was standardized/*z*-transformed among the tokens produced by the same speaker, so the standardized F0 assumed a value in  $(-\infty, \infty)$ . The analyses revealed that F0-normalization methods had a profound impact on the patterns uncovered by the statistical models: the models suggested that aspirated tokens had a *higher* scaled-F0 than unaspirated tokens, but aspirated tokens had a *lower* standardized-F0 than unaspirated ones. However, unpacking the mathematical differences behind the two normalization methods reveals that they reflect different properties of an F0 contour. While F0 scaling characterizes the steepness of F0 change within a token, F0 standardization captures the vowel-onset F0 of a token relative to all the other tokens. The combined results therefore indicated that, while aspirated stops had a lower vowel-onset F0 than unaspirated stops on average, aspirated stops tended to start with a higher F0 relative to the rest of their F0 contour than unaspirated ones. In addition, this study reiterates the importance of careful consideration of F0-normalization approaches for research in the domain of vowel-onset F0 perturbation. Given that this dissertation concerns the relative ordering of vowel-onset F0 following different consonant classes, F0 standardization (i.e., *z*-transformation) will be used in the next chapters.

## Chapter 3

# The Dual Role of Post-Stop F<sub>0</sub> in the Production and Perception of Stops in L1 Mandarin-L2 English Bilinguals

### 3.1 Introduction

Speech sounds contrast on a multitude of continuous acoustic dimensions, with some dimensions being used as primary cues to a phonological contrast while others play a more secondary part. Following Toscano and McMurray (2010), I use the term *cue* to refer to any source of information that allows the perceiver to distinguish between different responses (e.g., the response might be whether the sound is an [i] or an [a]). An example that is often given in this connection is Lisker's (1986) finding that potential cues to word-medial voicing in English (e.g., *rapid* vs. *rabid*) include duration of the preceding vowel, duration of the closure, voice onset time (VOT), presence of vocal fold vibration during closure, burst amplitude, fundamental frequency (F<sub>0</sub>) going into and out of the closure, among others. However, the reverse—that an acoustic dimension can serve as a cue for multiple phonological contrasts—is also true but often less studied. For instance, formant

frequency is not only an important cue for vowel quality, but the transition for a formant frequency band also cues the place of articulation for stop consonants (e.g., Liberman et al., 1954). Given this many-to-many mapping between phonological contrasts and acoustic dimensions, ambiguity naturally arises about how speakers encode various cues for a contrast and how listeners infer potential contrasts from a cue.

The current study explores this ambiguity from the perspectives of both speech production and perception. Specifically, I am interested in (i) whether and how F0 is used by speakers of a tonal language to signal and perceive both phonological *voicing* in stops and lexical tone simultaneously, and (ii) whether the use of F0 might be mediated by different language contexts. Bilingualism of a tonal and a non-tonal language offers a window into these inquiries. In this study, the two questions are addressed in tandem by comparing first language (L1) Mandarin-second language (L2) English bilinguals' performances in production and perception of Mandarin and English stops. The production task involves the participants reading aloud words with a stop in the onset position, while the perception part asks the participants to respond in a forced-choice identification task based on synthetic continua of both VOT and F0 values.

## 3.2 Background

### 3.2.1 Fundamental Frequency as a Cue to Lexical Tone

Similar to segments, lexical tones contrast on multiple acoustic dimensions, such as duration and intensity; however, F0 has long been established as the most important acoustic correlate for tonal distinctions, as far as Mandarin is concerned (Ohala, 1978). Indeed, the tone letters in the International Phonetic Alphabet are in their essence a discretized representation over a speaker's full pitch range, and the descriptions for lexical tones in Mandarin closely follow the F0 as it unfolds over a syllable—Tone 1: high-level ˥, Tone 2: mid-rising ˨˨˩, Tone 3: low-dipping ˨˩˨, and Tone 4: high-falling ˥˩. Even though F0 is not the only dimension that covaries with each tone in production (Ho, 1976), and it is not the only dimension that listeners take advantage of when distinguishing tones (e.g., Blicher et al., 1990), it is

the primary source that Mandarin users rely on to signal and extract information regarding tonal contrast (Gandour, 1978).

In this study, I restrict the scope to only Tone 1 and Tone 4 for both theoretical and practical considerations. On the theoretical side, Tone 1 and Tone 4 are the only two tones in Mandarin that start with the same phonological tonal register (i.e., both start with a high target), so listeners need to track the F0 trajectory, at least for the initial portion of a tonal contour initiated with a high register, to reliably tell these two tones apart. This is an important consideration for the design of the perception experiment, as will be explained in Section 3.2.2. Also, given that both Tone 1 and Tone 4 begin in the upper part of the pitch range, post-stop F0 behaviors, which will be discussed in the next section, should be more comparable across these two tones, as there is evidence suggesting that post-stop F0 is contingent on pitch height.

### **3.2.2 Fundamental Frequency as a Cue to Stop Voicing**

#### **Post-stop F0 in English**

It has been observed that F0 in the vowel following a stop consonant tends to correlate with voicing distinctions cross-linguistically (e.g., Cantonese [Francis et al., 2006; Luo, 2018; Ren and Mok, 2021], English [Hanson, 2009; Hombert, 1978; Hombert et al., 1979; House and Fairbanks, 1953; Lea, 1973; Lehiste and Peterson, 1961; Ohde, 1984], French [Kirby and Ladd, 2016], German [Kohler, 1982], Japanese [Gao and Arai, 2018], Korean [Han and Weitzman, 1970; Jun, 1996], Mandarin [Chen, 2011; Guo, 2020; Howie, 1976; Luo, 2018; Xu and Xu, 2003], Russian [Mohr, 1971], Spanish [Dmitrieva et al., 2015], Thai [Ewan, 1976; Gandour, 1974], Xhosa [Jessen and Roux, 2002], Yoruba [Hombert, 1978]). This phenomenon is commonly labeled as post-stop F0 perturbation, pitch skip, obstruent intrinsic F0, co-intrinsic pitch, or onset F0 perturbation. For English, whose six stops come in phonologically voiced-voiceless pairs: /b/-/p/, /d/-/t/, and /g/-/k/, it is well-established that F0 at vowel onset is significantly higher following phonologically voiceless stops than following phonologically voiced ones, regardless of the presence of actual vocal fold vibration (e.g., Abramson and Lisker, 1985; Dmitrieva et al., 2015). This type of patterning has led Kingston and Diehl (1994)

to argue that post-stop F0 is not purely a result of intrinsic physiological dependencies between the articulatory and/or aerodynamic properties of the production of degrees of prevoicing or voicing delay—instead, it is at least partially the result of controlled processes referring to the phonological status of the consonant series.

The perceptual consequences of post-stop F0 to the voicing contrast are also firmly established for English: a higher post-stop F0 tends to lead to more voiceless responses than a lower F0, especially when VOT is ambiguous (Francis et al., 2006; Whalen et al., 1990, 1993). Some authors have attributed the perceptual effects of post-stop F0 on voicing decisions to the observation that a low F0 enhances the perceptual “voicedness” of a stop by highlighting the percept of low-frequency periodic energy in the proximity of the stop release (Kingston and Diehl, 1994; Kingston et al., 2008).

### **Post-stop F0 in Mandarin**

As previous findings pertaining to the post-stop F0 perturbation effect in Mandarin have been reviewed in Chapter 2, the reader is referred to Section 2.1.1 for a detailed description. Suffice it here to highlight the main findings in Table 3.1.

More broadly, the issue of post-stop F0 perturbation in Mandarin is related to the debate of whether there is a trade-off between post-stop F0 and tone, and of whether the existence of tone attenuates the degree of post-stop F0 difference. While there are some studies that provide a positive answer (e.g., Gandour [1974] for Thai and Hombert [1978] for Yoruba), larger magnitudes have also been reported in tonal languages (e.g., Phuong [1981] for Northern Vietnamese, Shimizu [1994] for Thai, Xu and Xu [2003] for Mandarin, and Francis et al. [2006] for Cantonese). In the current study, the parallel production experiments in Mandarin as well as English allow us to address this debate from a bilingual perspective. That is, the production data in Mandarin and English enables a comparison of the degree of post-stop F0 difference across a tonal and a non-tonal language within the same speaker.

The perceptual contribution of post-stop F0 to the voicing contrast in Mandarin is substantially less studied. To my knowledge, Guo (2020) is the first to systematically study whether post-stop F0 is used by Mandarin speakers as a cue when

**Table 3.1:** Main findings from previous experimental studies on the post-stop F0 perturbation effect in Mandarin.

<b>Study</b>	<b>Findings</b>
Howie (1976)	- F0 after aspirated higher than F0 after unaspirated - F0 after unaspirated higher than F0 after sonorant
Xu and Xu (2003)	- F0 after aspirated <i>lower</i> than F0 after unaspirated - Magnitude of F0 difference varies by tone
Chen (2011)	- F0 after aspirated higher than F0 after unaspirated - Magnitude of F0 difference varies by tone - Tone effects vary by gender
Luo (2018)	- F0 after aspirated higher than F0 after unaspirated - F0 after unaspirated higher than F0 after sonorant - Magnitude of F0 difference varies by tone
Guo (2020)	- Direction of F0 perturbation varies by tone - F0 after aspirated higher than F0 after unaspirated in Tone 1 and Tone 4 - F0 after aspirated <i>lower</i> than F0 after unaspirated in Tone 2 and Tone 3 - F0 after unaspirated higher than F0 after sonorant across all tones

tasked to distinguish the stop voicing contrast in Mandarin. Using a two-alternative forced choice (2AFC) paradigm, Guo (2020) shows Mandarin speakers capitalize on post-stop F0 to decode consonantal voicing information. However, the identification experiment in her study only required the listener to distinguish aspirated versus unaspirated stops in the context of the same lexical tone (i.e., the two alternatives in the 2AFC paradigm only differed in stop voicing but shared the same lexical tone), and so it is still unclear whether Mandarin listeners continue to use post-stop F0 as a cue for voicing when they have to extract tonal information from pitch at the same time. The design of the current perception experiment addressed this problem, as explained in Section 3.4.3.

### 3.2.3 Post-Stop F0 at L1 Production-Perception Interface

While there is clear evidence that post-stop F0 functions as a cue for voicing in production as well as in perception *separately*, outcomes from attempts to link the cue use *across* the two modalities remain inconclusive. More generally, based on the proposal that perceptual cue weights arise from statistical regularities in the input (e.g., Francis et al., 2008; Holt and Lotto, 2006; Toscano and McMurray, 2010), one would anticipate the relative informativeness of a cue in a speaker's productions of a contrast to be predictive of the reliance assigned to that cue in perceiving the same contrast. Theories that posit a strong and/or direct connection between production and perception, such as Motor Theory (Lieberman and Mattingly, 1985), Direct Realism (Fowler, 1986), and exemplar models (Johnson, 1997; Pierrehumbert, 2001, 2003), also express such a view. However, although it is established that distributional patterns in production are exploited as cues in perception at the macro level, efforts to find correlations between use of the same cue across production and perception at the micro or individual level have been met with mixed success. For example, while Zellou (2017) found that individuals' production of anticipatory nasal coarticulation on vowels in English was correlated with their patterns of perceptual compensation, Kataoka (2011) found no significant correlation between Californians' production and perception of /u/-fronting in alveolar contexts. Zooming in on the use of post-stop F0, even as the use of post-stop F0 as a perceptual cue for stop voicing reflects the differential F0 at vowel onset in production on a population level, correlational analysis on an individual level has yet to reveal a more direct connection. For instance, the importance an English speaker assigns to post-stop F0 in production does not seem to predict the perceptual reliance of the same cue from the same individual (Shultz et al., 2012). A similar lack of relationship in post-stop F0 cue use for Spanish speakers is reported in Schertz et al. (2020). This study revisits this topic and explores whether there is a direct link between production and perception for the use of post-stop F0 in Mandarin, at both the population and individual levels.

### 3.2.4 Post-Stop F0 at L2 Production-Perception Interface

If producing and perceiving a phonological contrast means navigating between various acoustic dimensions, learning a phonological contrast in an L2 then involves adapting the weight associated with relevant dimension to approach that of monolingual L1 speakers of the L2 in question. Some work on L2 sound production and perception has put an emphasis on how L2 learners acquire foreign contrasts that rely primarily on dimensions that are not used in similar native contrasts. For instance, the difficulty for Japanese speakers to distinguish the English /r/-/l/ contrast is ascribed to the fact that this English contrast relies mainly on a difference in third formant values, whereas it is the second formant that Japanese speakers use to distinguish the categories (Iverson et al., 2003; Lotto et al., 2004; Miyawaki et al., 1975).

Another interesting line of research focuses on cases in which a first language (L1) contrast primarily relies on *more* cues than the corresponding L2 contrast. A study in this direction is Schertz et al.'s (2015) research on how L1 speakers of Korean, which uses both VOT and post-stop F0 as primary cues for its three-way stop distinction, produce and perceive the L2 English stop contrast, which relies primarily only on VOT.

The current work represents a case study that is in some sense sandwiched between the two threads of research discussed above. In particular, similar to English, Mandarin relies primarily on VOT to signal its stop voicing contrast; this therefore distinguishes the case of L1 Mandarin speakers learning the L2 English stop contrast from that of L1 Japanese speakers coping with the English /r/-/l/ contrast. However, the case study in this work also deviates from Schertz et al.'s (2015) study of L1 Korean speakers in that, unlike Korean, which uses *both* VOT and F0 for its three-way stop contrast, L1 Mandarin speakers' sensitivity to F0 stems from the lexical tones in the language's phonological inventory. Crucially, for L1 speakers of a tonal language learning a non-tonal L2, F0 is an ambiguous cue that signals tonal contrasts in L1 but non-tonal contrasts in L2. Examining this sort of scenario is therefore important for understanding to what extent L2 learners learn to transfer cues across phonological domains (i.e., using F0 as a suprasegmental cue to using it as a segmental cue) during L2 sound category acquisition.

In fact, the research questions raised here have been partially addressed by Guo (2020). In her study, she had a group of Mandarin-English bilinguals dominant in Mandarin produce a set of Mandarin and English words typifying stop voicings in the respective languages, and the same group of participants also took part in 2AFC perception experiments, identifying Mandarin and English words with different combinations of VOT and post-stop F0 values. Visual inspection of her production results suggests that the difference in post-stop F0 between long-lag stops and short-lag stops is smaller in Mandarin than in English, though no statistical models were used to test this observation. In perception, her results also suggest that Mandarin listeners use post-stop F0 as a cue for stop voicing in both L1 Mandarin and L2 English word identification tasks, but whether the extent with which they relied on post-stop F0 differed according to the language context was not analyzed. In this study, these caveats were addressed with a different experiment design.

Much like the link between production and perception in L1, the production-perception interface in L2 has turned out to be elusive, potentially due to more individual variability induced by more diverse L2 learning experiences. While at the broad level, the perception patterns often mirror production patterns, and vice versa, work looking for production-perception links with respect to individual cue weights has had limited luck finding correlation between the two modalities. For example, in studying L1 Korean learners' production and perception of the stop voicing contrast in English, Schertz et al. (2015) find considerable individual difference in L2 English perceptual categorization strategies in spite of the relative homogeneity of their L2 English production. In the current work, the L2 production-perception interface was also briefly examined, focusing on the use of post-stop F0 in L1 Mandarin learners' production and perception of English stops.

### **3.2.5 L1 Influence on L2 Cue Use**

Given that the target population in this study is L1 Mandarin-L2 English speakers, one would expect the usage patterns of multiple acoustic dimensions in their L2 English to be influenced by their L1 Mandarin. Such an L1-to-L2 influence can be understood in the frameworks of two major theories of L2 speech sound acquisition—the Revised Speech Learning Model (SLM-r, Flege and Bohn, 2021)

and the Perceptual Assimilation Model's extension to L2 acquisition (PAM-L2, Best and Tyler, 2007). Both models relate the patterns of L2 sound acquisition to L1 phonology by assuming that L2 sounds are assimilated to L1 sound categories whenever possible. The difficulty of L2 sound discriminability is therefore projected from the phonetic similarity between L1 and L2 sounds, and the patterns of assimilation from L2 to L1 categories. Given that both the English and Mandarin stop contrasts make use of VOT as the primary cue, that the absence/presence of aspiration is an important indicator for phonological voicing, and that both languages have two stop categories in terms of phonological voicing, English phonemically voiced (/b, d, g/) and voiceless (/p, t, k/) stops in the word-initial position will almost certainly be assimilated to Mandarin unaspirated (/p, t, k/) and aspirated stops (/p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>/) respectively. In the extreme case where English stops are processed as Mandarin stops, one would expect the participants to transfer their native Mandarin cue-weighting strategies to English in either perception, as predicted by the PAM-L2, or across production and perception, as predicted by the SLM-r.

However, more recent works have also demonstrated that late L2 learners are able to fine-tune the use of various acoustic dimensions in different language contexts. For instance, Amengual (2021) examined the VOT of the English, Japanese, and Spanish /k/ in the productions of L1 English-L2 Japanese bilinguals, L1 Japanese-L2 English bilinguals, and L1 Spanish-L2 English-L3 Japanese trilingual and found that all three groups of speakers produced language-specific VOT patterns for each language, despite evidence of cross-linguistic influence. In perception, Casillas and Simonet (2018) investigated whether English beginner learners of Spanish at the early stages of their development could manifest the double phonemic boundary effect in VOT—that is, whether these bilinguals shift the perceptual VOT boundary according to the language mode they are in—and found that they were indeed able to manifest the effect, suggesting that the ability of switching between language-specific perceptual modes can be acquired later in life, as the SLM-r would predict. It is therefore possible that the bilingual participants in this study are capable of adjusting the weight of post-stop F0 according to the language context. The production and perception experiments presented in this work allow for robust investigation of this possibility.

### 3.2.6 Goals of the Current Study

The use of F0 as a medium for the lexical tones in Mandarin provides an opportunity to examine whether F0 also functions as a cue for stop voicing in production—as has been found for a number of non-tonal languages—and as a cue for stop voicing in perception when Mandarin listeners also need to extract tonal information from F0. With respect to production, previous work has not converged to a definite conclusion, so the current study aims to first establish the post-stop F0 production patterns in the participating speakers. Note, however, that the conclusion drawn from this study might not generalize to monolingual Mandarin speakers or to speakers of a different Mandarin variety, given that the participants in the current study are all bilinguals and grew up in China. Concerning perception, while there is evidence that Mandarin listeners take advantage of post-stop F0 as a cue for stop voicing, the experiment with which this observation was made did not require the listeners to simultaneously track F0 for lexical tone, so it is therefore still an open question whether Mandarin listeners actually use post-stop F0 as a cue for stop voicing in more natural settings.

The second aim of this study is to investigate whether the use of post-stop F0 as a cue is sensitive to different language contexts. Capitalizing on the fact that the L1 Mandarin speakers that could be recruited in the university communities here were also L2 English speakers, one relevant question is whether Mandarin-English bilinguals use post-stop F0 cue to different extents, depending on the language “mode” they are operating in. If post-stop F0 is not solely due to physiological and/or aerodynamic reasons and is partially subject to active control, as postulated in Kingston and Diehl (1994), Mandarin-English bilinguals might actively, though subconsciously, suppress post-stop F0 in Mandarin because of the pressure to maintain tonal contours, which they do not have to do when speaking English. In perception, the demand to track F0 for lexical tone when perceiving Mandarin might prompt the bilingual listener to attribute variation in F0 partially to lexical tone, which makes them less likely to treat variation in post-stop F0 as an indicator for voicing. However, freed from the burden of tracking F0 for tone, as when they are perceiving English, the same listeners are now more likely to link the difference in post-stop F0 to consonantal voicing. These two scenarios could lead to bilinguals

using the post-stop F0 cue differentially across their two languages in both production and perception, which would be reflected as different cue weights for post-stop F0. On the other hand, given that the bilinguals are dominant in Mandarin, they may simply import their cue-weighting strategies for Mandarin to English, as predicted by the SLM and PAM-L2, resulting in the same weight for post-stop F0, regardless of language. The hypotheses and the corresponding predicted results just described are summarized in Table 3.2. The conducted production and perception experiments can help distinguish between the two possibilities.

An additional aspect that is foregrounded in this study is individual variability in participants' production and perception in their L1 and L2. Specifically, the relationship between individual participants' production and perception of post-stop F0 is explored. For this purpose, individual participants' production and perceptual post-stop F0 weights in their L1 and L2 are derived first. Correlation analyses are then used to examine whether individuals' post-stop F0 weights are statistically linked either within the same modality but across languages, or within the same language but across modalities.

### **3.3 Production Experiment**

This experiment examined non-early Mandarin-English bilinguals' productions of Mandarin and English word-initial stops and sonorants on vowel-onset F0.

#### **3.3.1 Participants**

A total of 103 English-Mandarin bilinguals dominant in Mandarin participated in the experiment, but only the data from a subset of 25 participants (14 female, 11 male;  $\text{Mean}_{\text{age}} = 20.9$  years,  $\text{SD}_{\text{age}} = 2.1$  years) were analyzed after excluding participants not meeting the inclusion criteria. The demographic information of the included participants is given in Table D.1 in the appendix. All participants were recruited from the linguistic participant pools at the University of British Columbia or the University of Toronto, and they received partial course credit for participation. For their production data to be considered in the analyses, a participant must satisfy all of the following criteria, which was determined based on their responses to the Language Background Questionnaire (Appendix A) and the

**Table 3.2:** Predicted production and perception results under difference hypotheses.

<b>Production</b>	
<b>Hypothesis</b>	<b>Predicted production results</b>
Post-stop F0 purely due to physiological / aerodynamic reasons (e.g., Kohler, 1984; Ladefoged, 1967; Ohala and Ohala, 1972) or total transfer of post-stop F0 cue use in Mandarin to English, as predicted by the SLM-r	Post-stop F0 difference the same in Mandarin and English tokens
Post-stop F0 partially subject to active control (Kingston and Diehl, 1994)	The extent of post-stop F0 difference might depend on the language (i.e., larger in English than in Mandarin)
<b>Perception</b>	
<b>Hypothesis</b>	<b>Predicted perception results</b>
Transfer of the Mandarin cue-weighting strategy to English, as predicted by the SLM-r and PAM-L2	Post-stop F0 weights the same across Mandarin and English
Flexibility in cue use: attributing variation in post-stop F0 partially to lexical tone and partially to stop voicing in Mandarin, but only to stop voicing in English	Post-stop F0 weights depend on the language context (i.e., a higher weight in English than in Mandarin)

Bilingual Language Profile Survey (Appendix B):

1. They completed all required experiment components;
2. They self-report as a native speaker of Mandarin;
3. They have at least one primary caretaker whose native language is Mandarin;
4. They are not simultaneous/early/childhood bilingual in Mandarin and English (i.e., they were exposed to English only after entering elementary school and did not receive their formal education in English prior to high school or university) but could have the local Chinese language as another early language;

5. They lived in China for at least 10 years between birth and age 15.

A number of additional inclusion guidelines, which are based on their audio recording quality and their performance in the perception experiment, were applied to further constrain the data entering the analyses. These inclusion guidelines are given in Section 3.3.5 and Section 3.4.4 respectively.

### **3.3.2 Stimuli**

This section describes the principles behind the selection of Mandarin and English production stimuli. The same logic was used for both languages, with adaptations to accommodate the phonotactic constraints of each language.

#### **Mandarin Stimuli**

The Mandarin stimuli consisted of 27 monosyllabic Mandarin words in isolation, as provided in Table 3.3. These words had onsets that exemplified the two laryngeal categories—voiceless aspirated and voiceless unaspirated—in Mandarin, as well as the sonorants /m/, /n/, and /l/. The sonorants were included to serve as the baseline against which the phonological voicing of stops was compared. To increase the generalizability of the findings, words with stops at three places of articulation (i.e., labial, alveolar, and velar), crossed with two levels of vowel heights (high: /i/, low: /a/, embedded in /ai/; /ai/, as opposed to /a/, was used because words with /ai/ are phonetically more similar to the English words used in the English production counterpart; see Section 3.3.2), were included. Given that lexical tone has been reported to modulate F0 perturbation in Mandarin (Guo, 2020), and that the influence of individual lexical tones is outside the scope of the current study, only Tone 1 and Tone 4 syllables were considered. Both tones start with a high pitch register and have been found to pattern together in conditioning post-stop F0 perturbation, making their production data more comparable to each other. Note also the existence of systematic and accidental gaps that prevented a fully crossed combination of the onsets, vowels, and tones. For instance, Mandarin disallows the occurrence of a velar stop before a high front vowel, so syllables such as \*/k<sup>h</sup>i/ and \*/ki/ are missing in Mandarin altogether. It is, however, accidental gaps in the language that caused \*/mai̯/, \*/ni̯/, etc., to be absent.

The stimuli were presented to the participants in simplified Chinese characters. Given that Mandarin has a large number of homophones that are nonetheless distinguished by different characters, each stimulus was represented with a common character so that all of them should be familiar to the participants, with the exception of *kai4* 愧, which is not a highly frequent character. To make sure that the participant knew the pronunciation of this character, its pinyin <kai4> was added to the right side of this character when presented to the participant. Care was also taken to ensure that different characters were as visually distinct as possible, to avoid the potential confound from visual priming across trials. For instance, while *pi1* could be represented with both 披 and 批, 披 was chosen because 批 shares the component 比 with another stimulus *pi4* 屁.

**Table 3.3:** Stimuli used in the Mandarin production experiment. The — symbol represents a consonant-vowel combination that violates the Mandarin phonotactics, and the = symbol stands for an accidental gap.

	/p <sup>h</sup> /	/p/	/t <sup>h</sup> /	/t/	/k <sup>h</sup> /	/k/	/m/	/n/	/l/
/iŋ/	<i>pi1</i> 披	<i>bi1</i> 逼	<i>ti1</i> 踢	<i>di1</i> 低	—	—	<i>mi1</i> 咪	=	=
/iŋ/	<i>pi4</i> 屁	<i>bi4</i> 闭	<i>ti4</i> 替	<i>di4</i> 地	—	—	<i>mi4</i> 密	<i>ni4</i> 逆	<i>li4</i> 利
/aiŋ/	<i>pai1</i> 拍	<i>bai1</i> 掰	<i>tai1</i> 胎	<i>dai1</i> 呆	<i>kai1</i> 开	<i>gai1</i> 该	=	=	=
/aiŋ/	<i>pai4</i> 派	<i>bai4</i> 败	<i>tai4</i> 泰	<i>dai4</i> 带	<i>kai4</i> 愧	<i>gai4</i> 盖	<i>mai4</i> 卖	<i>nai4</i> 奈	<i>lai4</i> 赖

### English Stimuli

The English stimuli consisted of 19 monosyllabic words, as given in Table 3.4. These words were selected following the same principles of stimulus section for the Mandarin tokens: the onsets typified voiceless stops, voiced stops, and sonorant at labial, alveolar, and velar places, while the vowels were either the front high vowel /i/ or the diphthong /ai/. When a simple combination of an onset and an open vowel did not correspond to a common English word, another common word with the same onset and nucleus but with an additional voiceless-stop coda was used as the alternative. Voiceless-stop codas, instead of other consonant classes, were used because they formed common English words. Also, for the syllable /di/, both the letter *D* and the word *deep* were used as stimuli to prevent loss of data for /di/ due

to the participant not producing /di/ upon seeing *D*.

**Table 3.4:** Stimuli used in the English production experiment.

	/p/	/b/	/t/	/d/	/k/	/g/	/m/	/n/	/l/
/i/	<i>pea</i>	<i>bee</i>	<i>tea</i>	<i>D/deep</i>	<i>key</i>	<i>geek</i>	<i>me</i>	<i>knee</i>	<i>Lee</i>
/aɪ/	<i>pie</i>	<i>buy</i>	<i>tie</i>	<i>die</i>	<i>kite</i>	<i>guy</i>	<i>my</i>	<i>night</i>	<i>lie</i>

### 3.3.3 Procedure

The procedure was identical for both the Mandarin and English versions of the experiment, and the order in which the two versions were administered was counter-balanced across participants. The entire experiment took place online in response to constraints on in-person data collection due to COVID, with the participant being instructed to complete the experiment on their own computer in a quiet place. They were encouraged to use an external microphone to keep the fidelity of audio recordings as high as possible, though they could still participate using the built-in microphone on their device.

The experiment was implemented in jsPsych, version 6.1.0 (de Leeuw, 2015). The experiment started with a microphone check to ensure that the input source was set correctly, and that the recording was clear. The experimental trials commenced after three practice trials that aimed to familiarize the participant with the recording interface and experimental flow. Each stimulus was repeated three times in three blocks respectively with a self-timed break between blocks. Stimuli were presented in a randomized order within each block. Each trial began with a plus sign at the center for 500 ms, and the recording was initiated automatically at the same time. The stimulus then appeared at the center, replacing the plus sign, and the participant was asked to read aloud the stimulus in a clear and natural manner. The trial ended with the participant clicking the “submit” button, which stopped the recording, uploaded the audio file to the server, and triggered the next trial. In the event where the participant did not click anything, the trial would terminate on its own after 10 s. The entire production experiment lasted about 15 mins.

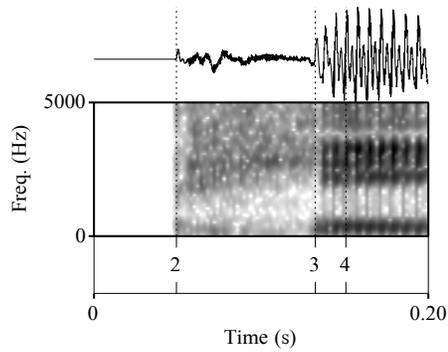
## Recording Annotations

All annotations and measurements were performed in Praat (Boersma and Weenink, 2021). The portion of the signal analyzed spanned from the beginning of the onset consonant to the end of the third pitch cycle of the nucleus vowel. The following guidelines were used when annotating tokens produced in either language, and examples of annotated tokens are displayed in Figure 3.1 and Figure 3.2:

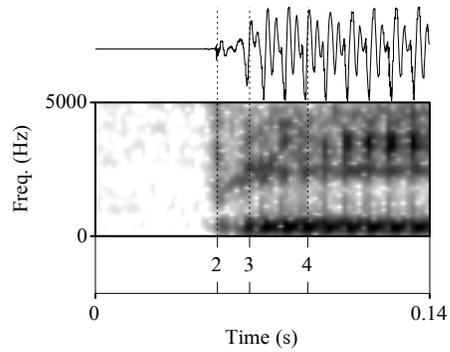
1. *Beginning of stop closure voicing*: In the cases where there was prevoicing for tokens with a voiced stop in English or, very rarely, with an unaspirated stop in Mandarin, all simple periodic chunks of the waveform before the release of the onset stop were marked as stop closure voicing.
2. *Beginning of stop burst*: For tokens with a stop onset, the beginning of the burst was marked at the starting point of perturbation in the waveform.
3. *Vowel onset*: The vowel onset was operationalized as the point where the (quasi) periodic part of the vowel first crossed zero in the positive direction.
4. *End of the third pitch cycle*: Following Cole et al. (2007) and Clayards (2018), the point marking the first 3 pitch cycles as counted from vowel onset was pinned in order to derive the onset F0.

### 3.3.4 Acoustic Measurements

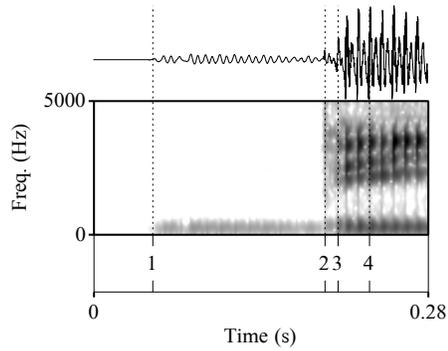
1. *Voice Onset Time (VOT)*: In line with the typical definition, VOT is defined as the time difference between the release of the stop and the onset of vocalic voicing. Accordingly, for prevoiced tokens (i.e., those with the beginning of stop closure voicing marked) VOT took a negative value, while VOT was positive for tokens where the onset of vocalic voicing followed the stop release. Tokens where the onset of vocalic voicing coincided with the stop release had a VOT of 0. To anticipate the analyses a bit, tokens with negative VOTs were included in the post-stop F0 models (Dmitrieva et al., 2015), but these tokens were excluded from the calculation of the post-stop F0 weight in production (see Section 3.3.7 for more detail).



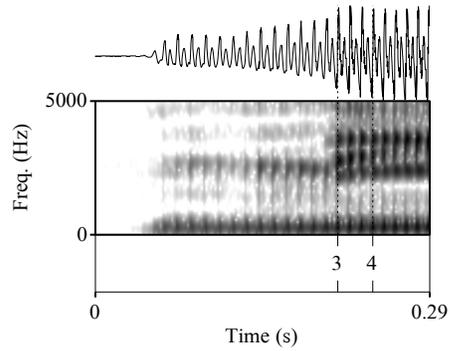
(a) *pea*



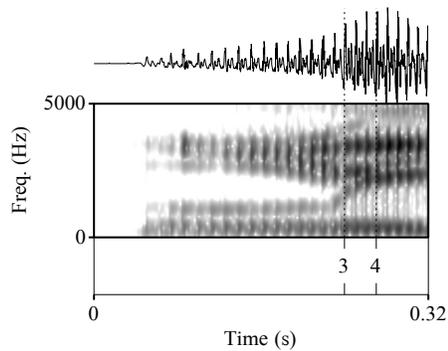
(b) *bee (short-lag)*



(c) *bee (lead)*



(d) *knee*



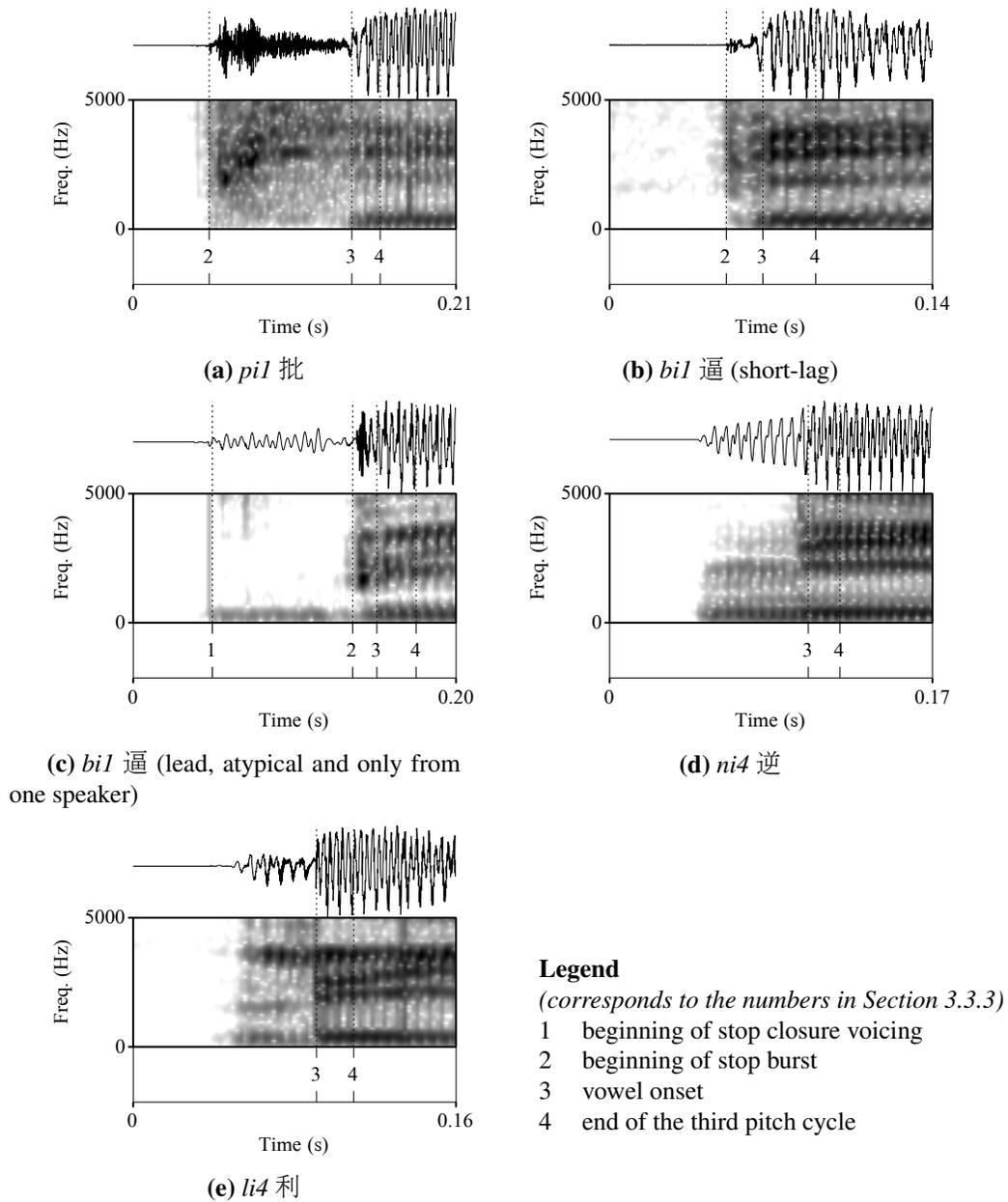
(e) *Lee*

**Legend**

(corresponds to the numbers in Section 3.3.3)

- 1 beginning of stop closure voicing
- 2 beginning of stop burst
- 3 vowel onset
- 4 end of the third pitch cycle

**Figure 3.1:** Examples of annotated English tokens, following the guidelines given in Section 3.3.3.



**Figure 3.2:** Examples of annotated Mandarin tokens, following the guidelines given in Section 3.3.3.

2. *Onset fundamental frequency (F0)*: This measurement was obtained by dividing 3 by the duration of the first 3 pitch cycles from vowel onset (i.e., 3 / (end of the third pitch cycle – vowel onset)). No F0-tracking algorithm was therefore involved for this measurement.

### 3.3.5 Participant Inclusion Criteria

Participants whose entire recordings (i) contained excessive background noise due to their doing the experiment in a noisy place ( $n = 1$ ), (ii) were extremely quiet that made it challenging to identify acoustic landmarks for annotation ( $n = 1$ ), or (iii) were of extremely low sampling rates ( $n = 1$ ), were omitted from the dataset altogether. There were also three participants who attempted the experiment more than once; in such a case, only the recordings from their first experiment attempt were considered. A subset of 25 participants was then selected based on their performance in the perception experiment, as explained in Section 3.4.4.

### 3.3.6 Omitted Data

The following types of tokens were excluded from all analyses: mispronunciations (11 Mandarin and 26 English), skipped tokens (2 Mandarin and 3 English), and technical issues (2 Mandarin and 4 English, including sporadic silent periods that overlapped with stop burst and/or vowel onset). Furthermore, tokens with creaky voice at vowel onset, subjectively determined by the author, were also omitted from all analyses for F0 estimation was unreliable for these tokens (50 Mandarin and 33 English). Altogether, 65 out of 2025 Mandarin tokens (3.2%) and 66 out of 1425 English tokens (4.6%) were excluded.

### 3.3.7 Statistical Analyses

The analyses consisted of two major parts: the first part addressed whether post-stop F0 had different values across the onset types in each language, and the second part focused on the quantification of production weight for post-stop F0 in each language. All models were fitted with Bayesian mixed-effects models, using `CmdStanR` (Gabry and Češnovar, 2021), an R interface for the Stan probabilistic programming languages (Carpenter et al., 2017). Bayesian models were chosen

because they return a distribution of potential values for all model parameters, making it more intuitive to assess the uncertainty associated with each parameter. In what follows, details about the statistical model employed are described.

### Post-Stop F0 Models

In this set of analyses, post-stop F0 was modeled as a Gaussian linear function of a number of variables that were properties of tokens or speakers. The names of predictor variables are given **boldface**, and different levels within a variable are indicated in SMALL CAPS.

*Variables* The dependent variable in all models was  $z$ -transformed post-stop F0. The decision to use  $z$ -transformation, as opposed to other normalization methods, such as semitone transformation, was justified in Section 2.5.2. In short, given that the research question concerns a token’s vowel-onset F0 as compared to the other tokens, normalization methods that do not flip the relative positions among tokens in terms of F0 should be used.  $Z$ -transformation is in turn preferred to semitone because the former has been found to have a better clustering property (Rose, 2016). The post-stop F0 values from both Mandarin and English production were  $z$ -transformed within each speaker. That is, a single  $z$ -transformation was applied to Mandarin and English production data together for each speaker.

Four token-level predictors were considered: the **voicing** of the onset consonant, **language/tone**, the **height** of the main vowel, and the **place of articulation (PoA)** of the onset consonant. Forward difference coding was used for **voicing** (ASPIRATED vs. UNASPIRATED and UNASPIRATED vs. SONORANT). Helmert coding was used for **language/tone** (ENG vs. mean of MAN T1 and MAN T4, and MAN T1 vs. MAN T4). Sum coding was used for **height** (HIGH, NON-HIGH = [1, -1]) and **PoA** (LABIAL, ALVEOLAR, VELAR, with LABIAL coded with -1). To account for how each predictor affected the realization of the voicing contrast, two-way interaction terms between **voicing** and all the other predictors were also included in the model comparison process. These first-order and second-order terms therefore constituted the population-level (“fixed-effect”) predictors.

For individual-level (“random-effect”) predictors, by-**speaker** effects consisted

of a random intercept and random slopes for all population-level predictors.

*Model Structure* Standardized post-stop F0 was modeled as a function of a subset of the predictor variables introduced above, using Bayesian linear mixed-effects models. All candidate models shared general specifications. Main-effect terms were included for the predictor variables selected in a particular candidate model. As mentioned above, two-way interaction terms being **voicing** and the other predictors were also considered. I did not, however, consider any three-way interactions as they are in general harder to interpret and could drastically slow down model sampling. All models also included by-speaker random intercepts, to account for variability in post-stop F0 of speakers beyond the effects of predictor variables. All possible by-speaker random slopes were also included to account for variability among speakers in the effects of predictors on post-stop F0 (Barr et al., 2013).

Each model was fitted with regularizing priors of Normal( $\mu = 0$ ,  $\sigma = 5$ ) for the intercept and all population-level parameters. An Exponential( $r = 1$ ) distribution was used as the prior for the error term as well as for the individual-level standard deviations. Correlations among individual-level effects used the Lewandowski-Kurowicka-Joe (LKJ) prior (Lewandowski et al., 2009) with  $\xi = 1$  (which is the default in the `brms` package [Bürkner, 2017]). All models showed no divergent transitions and had  $\hat{R}$  values close to 1 (i.e., all  $\hat{R} < 1.01$ ), which indicates that chains were well-mixed.

*Inference Criteria* The same inference criteria as those used in Chapter 2 was adopted here. To briefly recapitulate, evidence in each model was evaluated in two ways: (i) the posterior distributions of parameters, and (ii) comparison of models of different complexities. With regard to (i), I consider there to be strong evidence for a non-null effect if the 89% credible interval (CrI) for the parameter does not include 0. If the 89% CrI spans 0, but the probability of the parameter not changing direction is at least 89%, I consider this to represent weak evidence for a given effect. For (ii), model comparison was done by means of the Bayesian leave-one-out estimate of expected log pointwise predictive density (ELPD-LOO), which attempts to gauge a model's *predictive accuracy*. When the estimated absolute

difference in ELPD-LOO between two models is at least 4, and 0 is not within two standard errors of the estimated difference, there is evidence that the two models have different predictive powers. In the following sections, model parameters are reported in terms of marginal posterior means of parameters, 89% CrIs, and the probability of effect direction.

*Candidate Models* The construction of candidate models for model comparison relied both on prior knowledge about factors affecting post-stop F0 and on a compromise between model complexity and predictive accuracy. All the candidate models are given in Table 3.5. Note that, to save space, only fixed effects are listed in Table 3.5; however, as described in Section 3.3.7, all models also included the corresponding maximal by-speaker random structure. Accordingly, model selection involved evaluating the combined contribution of fixed and random effects (van Doorn et al., 2021). Given that vowel height is known to influence F0 (“intrinsic F0”, Whalen and Levitt, 1995) and that language and lexical tone can affect F0, the base model (i.e., M1) started with the factors **height** and **language/tone**. As one of the goals is to establish whether and how post-stop F0 might be influenced by phonological voicing, further models were constructed by incrementally adding terms that involved **voicing**. For example, the comparison between M1 and M2 assessed the contribution of voicing in predictive accuracy, and comparing M2 and M4 examined the importance of the interaction between voicing and vowel height in predicting post-stop F0 values. Furthermore, a model with **PoA** as a predictor (i.e., M3) also entered into comparison to confirm that place of articulation does not cause post-stop F0 to differ. Note that when a fixed effect was added into the model, the corresponding by-participant random slope was also added to the random-effects structure. The formal specification of the final model can be found in Section D.2.1 in the appendix.

**Table 3.5:** Candidate post-stop F0 models considered in model comparison, with their ELPD-LOO means and standard deviations. An intercept was included in each model but is omitted here to save space.

<b>Model</b>	<b>ELPD-LOO</b>	<b>ELPD-LOO standard error</b>	<b>Predictors</b>
M1	-3637.3	60.3	height + lang/tone
M2	-3221.8	67.3	height + lang/tone + voi
M3	-3215.5	67.3	height + lang/tone + voi + PoA
M4	-3205.5	68.2	height + lang/tone + voi + voi × height
M5	-3189.0	67.8	height + lang/tone + voi + voi × lang/tone
M6 (final)	-3173.4	68.7	height + lang/tone + voi + voi × height + voi × lang/tone
M7	-3174.3	69.2	height + lang/tone + voi + voi × height + voi × lang/tone + voi × height × lang/tone

### Post-Stop F0 Production Weight Model

The second set of analyses aimed to quantify the production weight associated with post-stop F0. A higher production weight means post-stop F0 is more reliable in separating different members of the contrast. Following Clayards (2018), the production weight was calculated based on the amount of overlap between the categories, which was quantified using Cohen's  $d$  (Cohen, 1988):

$$d = \frac{\mu_{\text{asp}} - \mu_{\text{unasp}}}{\sqrt{1/2(\sigma_{\text{asp}}^2 + \sigma_{\text{unasp}}^2)}},$$

where  $\mu_{\text{asp}}$  and  $\mu_{\text{unasp}}$  refer to the mean F0s of the aspirated and unaspirated categories respectively, and  $\sigma_{\text{asp}}^2$  and  $\sigma_{\text{unasp}}^2$  are the standard deviations of F0 of the aspirated and unaspirated categories respectively.

Cohen's  $d$  for post-stop F0 was calculated at the population level with all speakers as a whole and at the individual level for each speaker. Only tokens produced with a positive VOT were included in the calculation, as negative VOTs were rare in the data (i.e., 9 out of 1440 Mandarin stop-initial tokens from 1 speaker in Mandarin, and 40 out of 901 English stop-initial tokens from 5 speakers in English) and therefore were not representative of the norm of this speaker population. Additionally, rather than estimating cue weights from empirical data as in most previous work (e.g., Clayards, 2018; Schertz et al., 2015; Shultz et al., 2012), a statistical model was used to derive the weight, which allowed for uncertainty around the weight to be incorporated. For this purpose, a Bayesian mixed model was first fitted to obtain the means and standard deviations of F0 of the aspirated and unaspirated categories for the whole group and for each speaker. The model included a cross-category correlation structure and used partial pooling to estimate individual means and standard deviations. For instance, a speaker's mean post-stop F0 for the aspirated category was correlated with their mean post-stop F0 for the unaspirated category, and both mean values were informed not only by the speaker's own production data, but also by other speakers' data thanks to partial pooling. The estimated means and standard deviations were then fed to the Cohen's  $d$  formula above to derive the posterior distribution of the production weight within the model. As such, the post-stop F0 weights of the entire group and for each speaker were not

just a single numerical value but a *distribution* that also carried information about uncertainty.

### 3.3.8 Results: Production of Post-Stop F0

Mean production values and standard deviations for L1 Mandarin and L2 English stops and sonorants on VOT and post-stop F0 are given in Table 3.6. Distributions of standardized post-stop F0 values are plotted in Figure 3.3. ELPD-LOO means and standard errors for the candidate models are listed in Table 3.5. A higher ELPD-LOO value means the model has a better predictive accuracy, so, for example, M2 makes better predictions than M1. Finally, model comparison results are summarized in Table 3.7 in terms of difference in ELPD-LOO values and associated standard errors. Note that the difference score in each cell was computed by subtracting the ELPD-LOO value of the model represented in the column from the ELPD-LOO value of the model indicated in the row. For instance, the difference  $-415.5$  came from  $ELPD-LOO_{M1} - ELPD-LOO_{M2} = (-3637.3) - (-3221.8)$ .

The results of model comparison indeed confirmed the importance of phonological voicing in conditioning post-stop F0 (i.e., M1 vs. M2) and spoke to the importance of interaction between voicing and vowel height (i.e., M2 vs. M4), and between voicing and language/tone (i.e., M2 vs. M5). Place of articulation, however, did not seem to influence post-stop F0 (i.e., M2 vs. M3). Since no significant gain in prediction was observed past M6, M6 was selected as the best balance between model complexity and predictive performance among the models being compared. The interpretation and discussion presented below are therefore based on this model.

In presenting the results, summary statistics and visualizations derived from raw data are given first, followed by the output from the final model in terms of posterior distributions for key parameters. I first interpret population-level parameter estimates before moving on to individual-level estimates.

#### Population Results

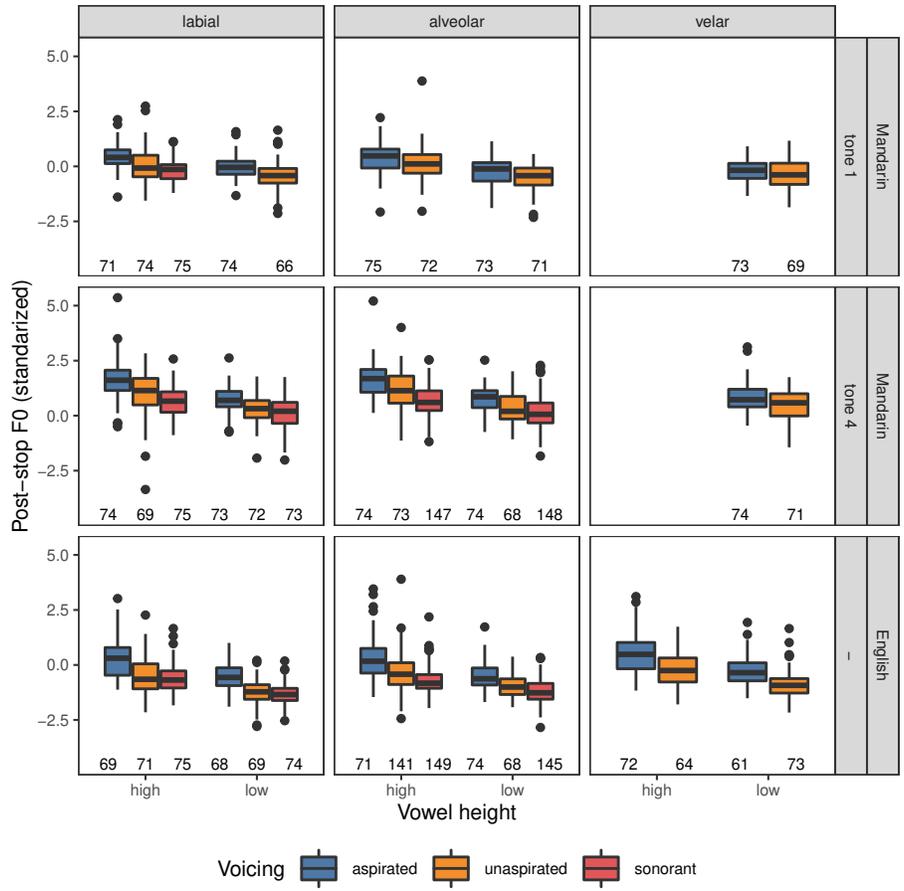
The marginal posterior distributions for population-level parameters from M6 are summarized in Table 3.8. As expected, both vowel height and language/tone con-

**Table 3.6:** Means and standard deviations for VOT and post-stop F0 in hertz (split by gender) for the 25 L1 Mandarin-L2 English bilinguals' productions of Mandarin and English word-initial stops and sonorants.

	<b>Mandarin</b>		
	<b>Aspirated</b>	<b>Unaspirated</b>	<b>Sonorant</b>
VOT (ms)	118 (32) <i>n</i> = 735	15 (18) <i>n</i> = 705	— —
F0, male (Hz)	158 (25) <i>n</i> = 323	150 (25) <i>n</i> = 328	152 (26) <i>n</i> = 227
F0, female (Hz)	286 (32) <i>n</i> = 412	277 (34) <i>n</i> = 377	280 (33) <i>n</i> = 291
	<b>English</b>		
	<b>Aspirated</b>	<b>Unaspirated</b>	<b>Sonorant</b>
VOT (ms)	111 (32) <i>n</i> = 415	7 (50) <i>n</i> = 486	— —
F0, male (Hz)	146 (20) <i>n</i> = 178	134 (19) <i>n</i> = 218	131 (19) <i>n</i> = 193
F0, female (Hz)	265 (28) <i>n</i> = 237	248 (30) <i>n</i> = 268	237 (26) <i>n</i> = 250

**Table 3.7:** Model comparison results for key model pairs, expressed in differences in ELPD-LOO and associated standard errors. Pairs that are judged to differ in predictive power are marked by asterisks.

<b>Model</b>	<b>M2</b>	<b>M3</b>	<b>M4</b>	<b>M5</b>	<b>M6</b>	<b>M7</b>
<b>M1</b>	−415.5* (29.5)					
<b>M2</b>		−6.3 (4.7)	−16.2* (7.5)	−32.8* (8.8)		
<b>M4</b>					−32.1* (8.7)	
<b>M5</b>					−15.6* (7.0)	
<b>M6</b>						.9 (4.5)



**Figure 3.3:** Standardized post-stop F0 values, normed by speaker, as a function of place of articulation, language/tone, vowel height, and phonological voicing. The number under each box represents the number of tokens in the distribution.

tribute to difference in post-stop F0. Specifically, the high vowel /i/ led to a higher onset F0 (HIGH – mean height:  $\bar{\beta} = .32$ , 89% CrI = [.27, .36],  $p(\beta > 0) = 1.00$ ), and Tone 4 tended to have a higher onset F0 than Tone 1 (MAN T1 – MAN T4:  $\bar{\beta} = -.91$ , 89% CrI = [-1.03, -.79],  $p(\beta < 0) = 1.00$ ). Also, participants’ L2 English tended to have a lower onset F0, in comparison with their L1 Mandarin (ENG – (MAN T1 + MAN T4)/2:  $\bar{\beta} = -.84$ , 89% CrI = [-1.02, -.66],  $p(\beta < 0) = 1.00$ ), which agrees with the general finding from the literature (Keating and Kuo, 2012; Lee and Sidtis, 2017). Critically, in both languages, aspirated stops had a higher post-stop F0 than unaspirated stops (ASP – UNASP:  $\bar{\beta} = .49$ , 89% CrI = [.41, .56],  $p(\beta > 0) = 1.00$ ), which in turn had a higher post-stop F0 than sonorants (UNASP – SON:  $\bar{\beta} = .29$ , 89% CrI = [.20, .39],  $p(\beta > 0) = 1.00$ ). In addition, the extent of post-stop F0 difference due to aspiration was contingent on language and tone as well, such that bilingual speakers’ English tokens showed an even bigger difference than Mandarin tokens ([ASP – UNASP]  $\times$  [ENG – (MAN T1 + MAN T4)/2]:  $\bar{\beta} = .25$ , 89% CrI = [.10, .39],  $p(\beta > 0) = 1.00$ ), and so did their Mandarin Tone 4 tokens in comparison with Tone 1 tokens ([ASP – UNASP]  $\times$  [MAN T1 – MAN T4]:  $\bar{\beta} = -.16$ , 89% CrI = [-.28, -.05],  $p(\beta < 0) = .99$ ).

### Individual Results

The distributions for key parameters involving voicing for each participant are visualized in Figure 3.4, and the individual posterior summaries for the full set of parameters are given in Table D.2 and Table D.3 in the appendix.

In both their Mandarin and English productions, there is strong evidence that all speakers produced a higher post-stop F0 following an aspirated stop than an unaspirated stop, as the 89% CrI is above 0 for all speakers in the [ASP – UNASP] panel in Figure 3.4. The [UNASP – SON] panel indicates that there is evidence that the onset F0 was higher adjacent to an unaspirated stop than adjacent to a sonorant, as the posterior means for all speakers are above 0. In addition, as evaluated by their 89% CrIs not spanning 0, the model is confident that this trend is strong for the majority of speakers. In terms of the post-stop F0 difference due to aspiration, some speakers evidently agree with the population pattern in having a bigger F0 difference in English, as indicated by their positive 89% CrIs in the [(ASP –

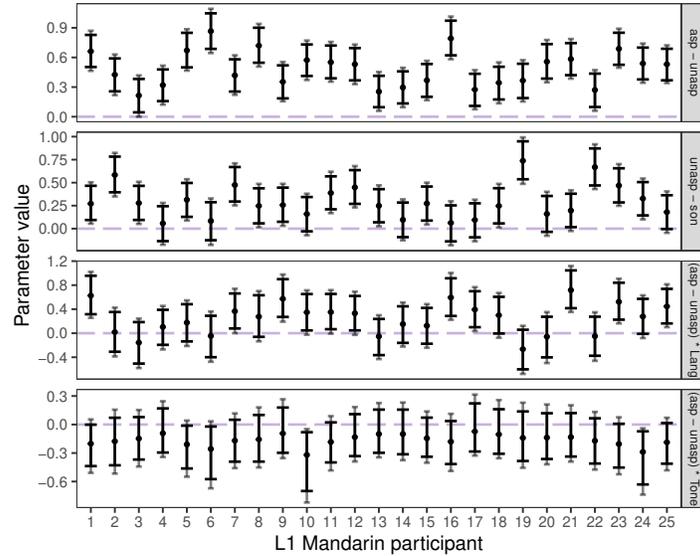
**Table 3.8:** Marginal posterior summaries for key population-level parameters from M6. The contrast coding scheme for each variable is described in Section 3.3.7. The parameters whose effects are judged to be strong are marked with \*\*, and those whose effects are judged to be weak are marked with \*.

Parameter	Mean	SD	89% CrI	$p(\text{dir})$
intercept	.01	.01	[−.01, .04]	$p(\beta > 0) = .84$
HIGH − (HIGH + LOW)/2**	.32	.03	[.27, .36]	$p(\beta > 0) = 1.00$
ENG − (MAN T1 + MAN T4)/2**	−.84	.11	[−1.02, −.66]	$p(\beta < 0) = 1.00$
MAN T1 − MAN T4**	−.91	.07	[−1.03, −.79]	$p(\beta < 0) = 1.00$
ASP − UNASP**	.49	.05	[.41, .56]	$p(\beta > 0) = 1.00$
UNASP − SON**	.29	.06	[.20, .39]	$p(\beta > 0) = 1.00$
[ASP − UNASP] × [HIGH − (HIGH + LOW)/2]*	.05	.04	[−.01, .12]	$p(\beta > 0) = .91$
[UNASP − SON] × [HIGH − (HIGH + LOW)/2]	.04	.04	[−.02, .10]	$p(\beta > 0) = .86$
[ASP − UNASP] × [ENG − (MAN T1 + MAN T4)/2]**	.25	.09	[.10, .39]	$p(\beta > 0) = 1.00$
[ASP − UNASP] × [MAN T1 − MAN T4]**	−.16	.07	[−.28, −.05]	$p(\beta < 0) = .99$
[UNASP − SON] × [ENG − (MAN T1 + MAN T4)/2]	−.03	.08	[−.15, .09]	$p(\beta < 0) = .64$
[UNASP − SON] × [MAN T1 − MAN T4]	.00	.10	[−.16, .15]	$p(\beta < 0) = .52$

UNASP) \* LANG] panel. For the other speakers, there does not seem to be a consistent trend, as even the posterior means are going in different directions. Finally, as shown in the [(ASP − UNASP) \* TONE] panel, even though the model cannot state confidently whether a specific speaker conformed to the population pattern (i.e., Tone 4 displayed a more differentiated post-stop F0 distinction between aspirated and unaspirated stops), there is clear evidence that many speakers trend in this direction, based on their posterior means.

### 3.3.9 Results: Production Weights of Post-Stop F0

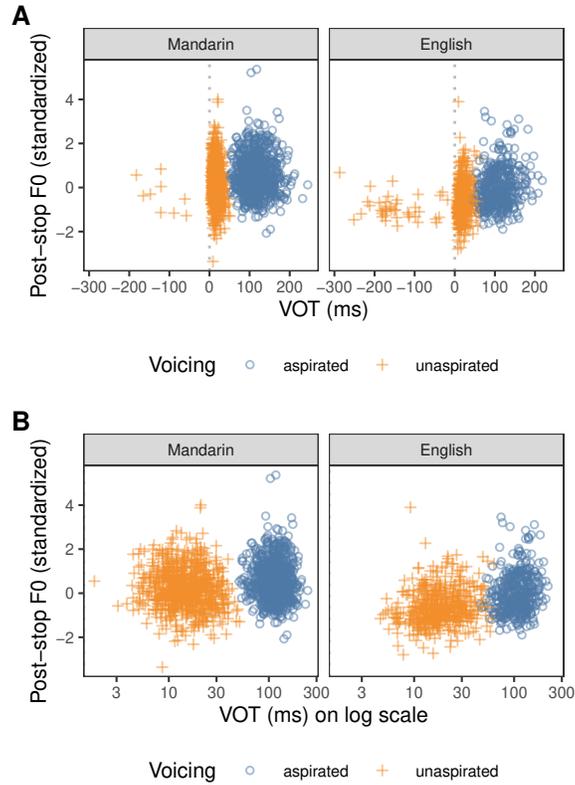
Standardized post-stop F0 values are plotted against raw VOT values for participants' Mandarin and English productions in Figure 3.5, and the distributions of production VOT and post-stop F0 weights, expressed in terms of Cohen's  $d$ , at the population level are graphed in Figure 3.6. Although the focus on this study is on the post-stop F0 cue, for completeness, the results for the VOT weight are also reported below.



**Figure 3.4:** Marginal posterior summaries for key parameters involving voicing for each individual speaker. The [asp – unas] panel shows the difference in F0 between aspirated and unaspirated stops. The [unas – son] panel shows the difference in F0 between unaspirated stops and sonorants. The [(asp – unas) \* Eng] panel shows the further difference in F0 between aspirated and unaspirated stops in English, in comparison to Mandarin. The [(asp – unas) \* Man T1] panel shows the further difference in F0 between aspirated and unaspirated stops in Mandarin Tone 1 tokens, when compared to Tone 4 tokens. The dots denote the posterior means. The inner error bars represent 89% CrIs, and the outer error bars represent 95% CrIs.

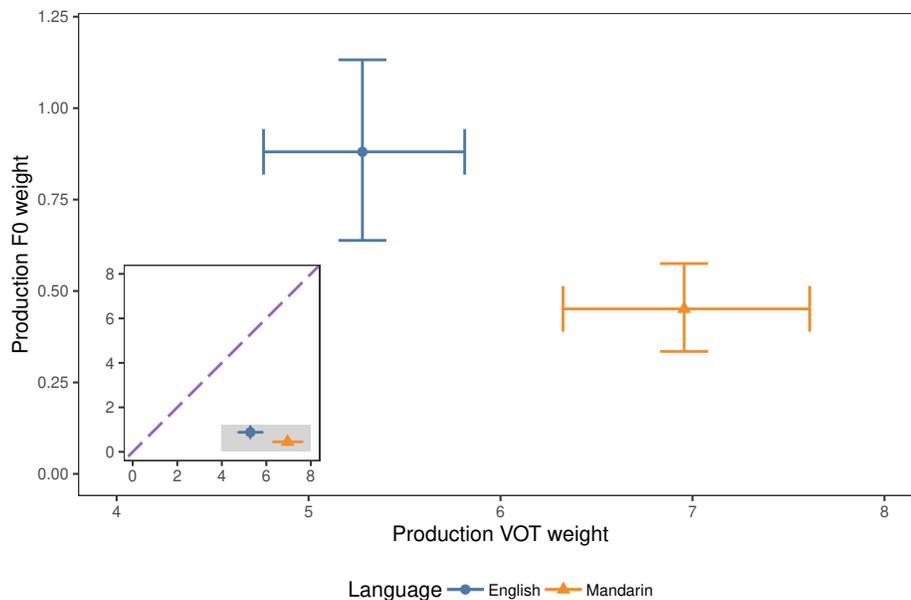
### Population Results

As can be seen in Figure 3.6, speakers as a group had a much higher weight for VOT than for post-stop F0, in both their Mandarin and English production. Also, regardless of language, there was more uncertainty surrounding the post-stop F0 weight than the VOT weight, as measured by the coefficient of variation (CV), which is defined as the ratio of the standard deviation to the mean (English:  $CV_{VOT} = .06$ ,  $CV_{F0} = .18$ ; Mandarin:  $CV_{VOT} = .06$ ,  $CV_{F0} = .17$ ). Contrasting the weights along the same dimension across languages, more weight was assigned to VOT in



**Figure 3.5:** Production values for VOT and post-stop F0 in the Mandarin and English stop contrasts. Raw VOT values are plotted along the  $x$ -axis, while standardized post-stop F0 values by speaker are plotted along the  $y$ -axis. **A.** Raw VOT values on a linear  $x$ -axis. **B.** Raw positive VOT values on a log  $x$ -axis to better show the category structure.

the Mandarin production ( $\bar{d} = 6.96$ , 89% CrI = [6.32, 7.61]), as compared to the English production ( $\bar{d} = 5.28$ , 89% CrI = [4.76, 5.81]), while the converse was true for the post-stop F0 weight: English tokens showed a heavier reliance on post-stop F0 ( $\bar{d} = .88$ , 89% CrI = [.64, 1.13]) than Mandarin tokens ( $\bar{d} = .45$ , 89% CrI = [.33, .58]).



**Figure 3.6:** Group-level production weights for VOT and post-stop F0 in each language. The embedded plot shows the same data but with the same scale along both axis, and the dashed line is  $y = x$ , which represents equal production weights for both dimensions. The shaded area indicates the part enlarged in the main plot.

### Individual Results

The reliability of each dimension for individual speakers, as estimated by Cohen’s  $d$ , is plotted in Figure 3.7. Conforming to the population pattern, all speakers assigned more weight to VOT than post-stop F0 in both their Mandarin and English productions (Figure 3.7A). When correlating weights along the two dimensions within language, no specific correlation pattern was discernible (see Figure 3.7B; Mandarin:  $\bar{\rho} = -.11$ , 89% CrI =  $[-.35, .15]$ ,  $p(\rho < 0) = .76$ ; English:  $\bar{\rho} = -.02$ , 89% CrI =  $[-.26, .21]$ ,  $p(\rho < 0) = .56$ ). However, when the VOT weights were correlated across languages, a strong positive correlation was observed ( $\bar{\rho} = .62$ , 89% CrI =  $[.40, .81]$ ,  $p(\rho > 0) = 1.00$ ), indicating that speakers who showed a larger VOT weight in Mandarin also tended to have a larger VOT weight in English (Figure 3.7C), replicating Chodroff and Baese-Berk (2019) and Johnson (2021). In

addition, for all but one speaker, VOT had more weight in their Mandarin tokens than their English tokens. For the post-stop F0 weight, most individuals (19 out of 25) echoed the population pattern in shifting their F0 weight upward when producing English tokens (Figure 3.7D), although there was no correlation in this cue across languages ( $\bar{\rho} = -.05$ , 89% CrI =  $[-.49, .40]$ ,  $p(\rho < 0) = .58$ ). Also notice that there was more individual variation for the post-stop F0 weight in the English production than in the Mandarin production, as indicated by a wider spread of individual weights in English than in Mandarin.

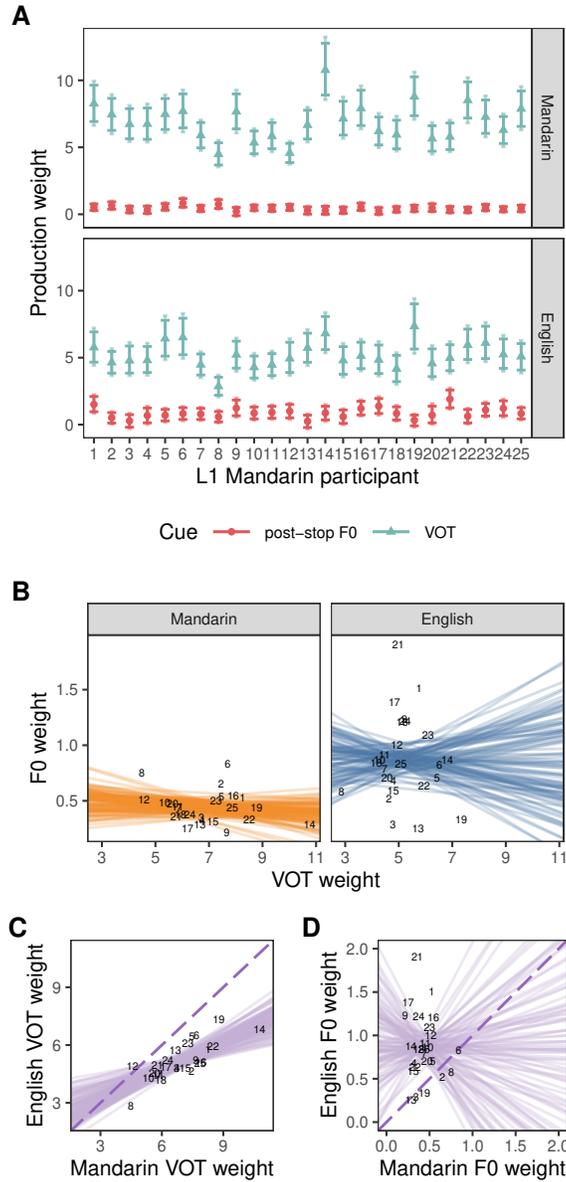
### 3.3.10 Interim Discussion: Production

The Mandarin production results reported here are in line with the recent work by Guo (2020) in terms of post-stop F0: both at the population and individual levels, the vowel-onset F0 following aspirated stops was higher than that following unaspirated stops. In addition, for most speakers, vowel-onset F0 after unaspirated stops was in turn higher than that after sonorants. Similar to their Mandarin production, the participants' English production also demonstrated a difference in post-stop F0 between aspirated and unaspirated series, but with an even larger F0 gap, both for the speakers as a whole and for over half of the individual speakers. This pattern again agrees with what has been found in Guo (2020).

Regarding cue weighting, VOT was the most reliable dimension distinguishing aspirated from unaspirated stops in both Mandarin and English, though it seemed that VOT assumed an even higher weight in Mandarin for almost all speakers (as measured by the posterior mean). The opposite pattern was observed for the post-stop F0 weight: English induced a higher weighting in this cue for most speakers. When the weighting between the two cues was correlated within each language, however, neither an enhancing nor a trading relationship was obtained.

## 3.4 Perception Experiment

The perception experiment turns to the perception of the Mandarin and English stop contrasts in the word-initial position by the same L1 Mandarin-L2 English bilinguals. The focus is on the contribution of post-stop F0 to categorization of the contrasts.



**Figure 3.7:** **A.** Individual speakers' production weights for VOT and post-stop F0. The posterior means are represented by the dots. The 89% CrIs are represented by the inner error bars, and the 95% CrIs are represented by the outer error bars. **B.** Post-stop F0 weights against VOT weights, separately for each language. The lines in the background are 100 lines of linear regression (i.e., post-stop F0 weight  $\sim$  VOT weight) fitted using 100 random posterior draws, separated for each language. **C.** Production VOT weights across languages. **D.** Production post-stop F0 weights across languages. In **C** and **D**, the solid lines represent 100 regression lines fit with 100 posterior draws, to show direction and uncertainty in the correlation. The dashed line is  $y = x$ , where the VOT weight for Mandarin equals that for English.

### 3.4.1 Participants

The same group of participants from the production experiment also took part in the perception experiment. The perceptual data analyzed here came from the same 25 participants whose production tokens were analyzed in the production experiment.

### 3.4.2 Stimuli

All stimuli were created from natural productions of the Mandarin words *bi1*, *pi1*, *bi4*, *pi4*, *yi1*, *mi1*, *mi4*, and *ni4* read by a 24-year-old male English-Mandarin speaker who speaks English as L1 but is also fluent in Mandarin. The prompts for production were words in isolation, which were presented three times to the model speaker in a randomized order. The recording was made on the Sound Devices MixPre-D audio mixer with a headset microphone. The produced syllables were then scrutinized by the author, and one token that was clear and did not have creaky quality was selected for each word as the raw tokens for manipulation.

#### Mandarin Stimuli

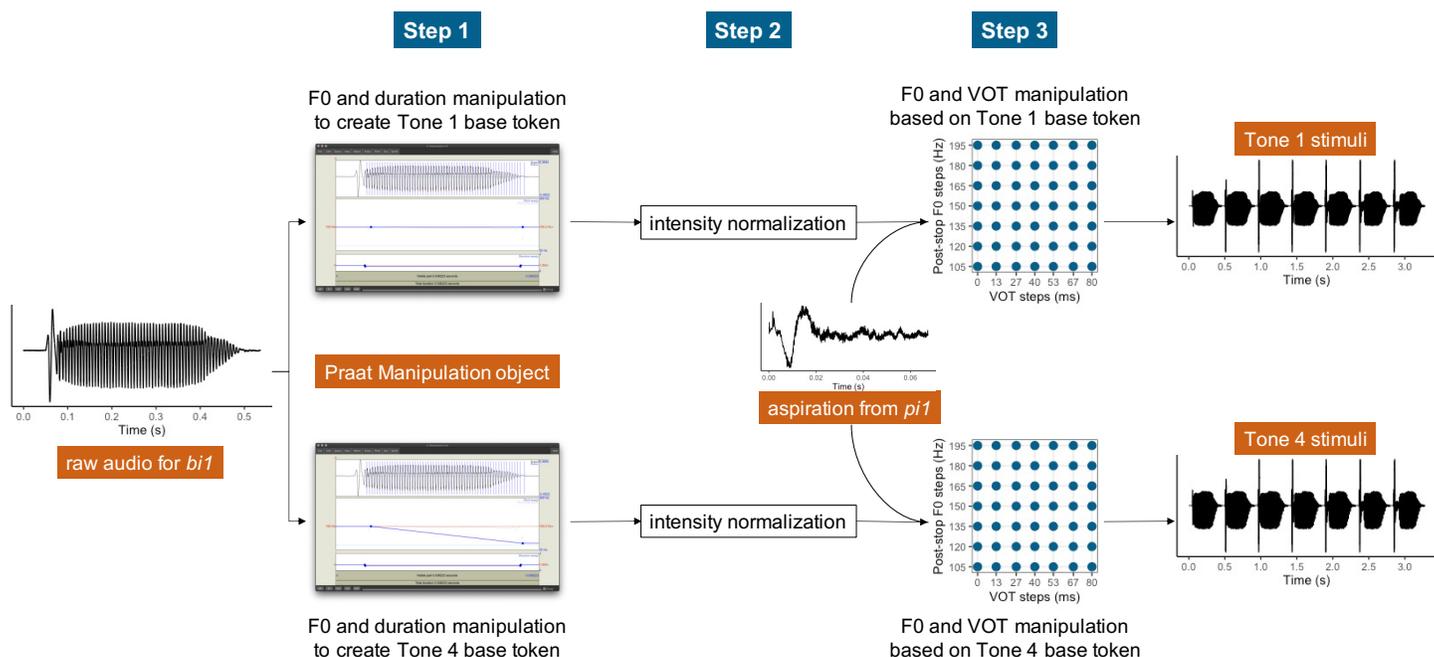
Stimuli could be categorized into the target or filler sets, with both sets containing Tone 1 and Tone 4 syllables. The target set was composed of syllables with a bilabial stop as the onset and the high vowel [i] as the nucleus, with the VOT of the stop and the initial F0 contour of the vowel manipulated. The manipulation along the VOT and F0 dimensions is summarized in Figure 3.9 and explained in the following paragraphs. Bilabial stops were used because they do not have lingual targets and therefore are expected to be coarticulated to a lesser degree with the following vowel (Schertz et al., 2020). The vowel [i] was selected because its formants are more stable across time in general (Hillenbrand et al., 1995).<sup>1</sup> In addition, the combination of bilabial stops with the high vowel also led to valid English lexical items *pea* and *bee*; this was critical given that the exact same stimuli

---

<sup>1</sup>The vowel [i] was also preferred from the perspective of VOT manipulation. Given that the starting values of the formant frequencies in the voiced part of the vowel could be substantially different depending on VOT, stimuli whose VOT values are manipulated with a “progressive cutback and replacement” approach (which was also used in this study) can have initial formant frequencies being correlated with VOT, leading formant cues to be a confound. Winn (2020) argues that since F1 of [i] is already low, the upward F1 transition common to the other vowels would be minimized, thus offering no covarying cue for VOT.

were used in the English version of the experiment as well. For fillers, Mandarin words *yi1*, *yi4*, *mi1*, and *mi4* were selected because they typified other onset types than the stop.

The target syllables were created by cross-splicing the vocalic portion of the *bi1* token and the burst+aspiration portion of the *pi1* token. The detailed steps of stimulus manipulation are described below.



**Figure 3.8:** Schematic diagram of stimulus-creating steps for target syllables.

The raw recording for *bi1* was first passed as Praat Manipulation objects to adjust F0 contours and vowel duration so that the resynthesized audios resembled Mandarin *bi1* and *bi4* in the citation form. The intensity of the resynthesized audios were then normalized to 75 dB before being further manipulated to create the full stimulus set.

**Step 1: Base tokens.** This step involved creating a Tone 1 and a Tone 4 base token for downstream manipulation. A Praat (Boersma and Weenink, 2021) Manipulation object was first created from the raw *bi1* recording. Then the vowel duration was equalized to 350 ms, which is approxi-

mately the mean duration of 416.2 ms for citation Tone 1 syllables and 307.8 ms for citation Tone 4 syllables (Yang et al., 2017).<sup>2</sup> The vowel duration was shifted to an ambiguous value to discourage the participant to use it as an additional cue for tone identification (e.g., Blicher et al., 1990). F0 trajectories were also manipulated to mimic natural Tone 1 and Tone 4 contours. For Tone 1, a simple pitch stylization was applied by setting both the initial and final F0 on the vowel to 150 Hz, so Tone 1 token was now literally a “level” tone. The F0 was set to 150 Hz because this was very close to the natural Tone 1 F0 register of the model speaker. Tone 4 was stylized as a linear F0 decline from 150 Hz to 60 Hz. The initial 150 Hz was to match the initial F0 value for Tone 1 while the final 60 Hz was set based on the model speaker’s natural Tone 4 production. The decision to recreate Tone 4 F0 contour from a Tone 1 item, instead of using a natural Tone 4 item, was to make sure that the same intensity profile was shared and would not be a confound.<sup>3</sup> After all the relevant parameters were set, the two base tokens were resynthesized with the Pitch Synchronous Overlap and Add (PSOLA) algorithm (Moulines and Charpentier, 1990) as implemented in Praat.

**Step 2: Intensity normalization.** The intensity of the two base tokens was scaled to 75 dB based on root-mean-square (RMS) amplitude, with a Praat script originally written by Matthew B. Winn.<sup>4</sup> The level 75 dB was chosen because this was approximately the intensity of the raw recording. Intensity normalization was done at this step, as opposed to at a later point when actual stimuli were synthesized, because Winn (2020) cautions that “the inclusion of a lengthy aspiration portion will justifiably reduce overall RMS intensity, so equalization would result

---

<sup>2</sup>These measurements are based on production of isolated monosyllables by 121 speakers (46 male and 75 female). Note that even though there seems to be a 100-ms difference between Tone 1 and Tone4, both tones have a standard deviation of about 90 ms in the syllable duration measurement, suggesting that the two tones overlap to a large extent in terms of their duration distributions.

<sup>3</sup>I have also attempted to create base tokens in the opposite direction: creating a Tone 1 item from a Tone 4 item. However, the resulting audio was noticeably unnatural, especially in the later portion where F0 needed to be raised from a low target of Tone 4 to a high target of Tone 1.

<sup>4</sup>[https://github.com/ListenLab/Praat/blob/master/Scale\\_intensity\\_all\\_sounds\\_in\\_folder.v1.txt](https://github.com/ListenLab/Praat/blob/master/Scale_intensity_all_sounds_in_folder.v1.txt)

in unnatural amplification of the syllable with voiceless onset” (p. 859). He therefore suggests that intensity amplification/attenuation should be applied before initiating VOT manipulation.

**Step 3: VOT and F0 manipulation.** The two intensity-equalized tokens were then modified, using another Praat script by Matthew B. Winn, to create word-initial bilabial varying in VOT duration and F0 at vowel onset.<sup>5</sup> The script relies on the Praat Manipulation object to perform manipulation and uses the PSOLA algorithm to resynthesize audio files. The reader is referred to Winn (2020) for detail concerning the script. The exact VOT and F0 values used and the duration for which F0 contour was altered are summarized in Figure 3.9 and explained in detail below.

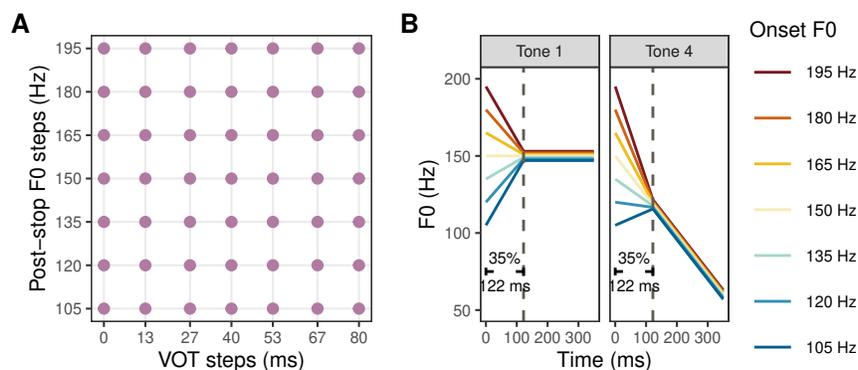
*VOT*: The duration of VOT in the base tokens was manipulated on a 7-step series ranging from 0 to 80 ms. The range endpoints were meant to span the VOTs of both English and Mandarin word-initial bilabial stops while still having enough resolution. Note that negative VOT was not in the manipulated range partially because “voiced” stops in word-initial position in English are very often realized as a short-lag stop with positive VOT (Fulop and Scott, 2021) and partially because including negative values would decrease the manipulation resolution. VOT was manipulated with a progressive-cutback-and-replacement approach—that is, the vowel onset of a word with a short-lag stop sound is progressively deleted and replaced with a roughly equivalent amount of the aspiration from its voiceless-onset counterpart (Winn, 2020)—to accommodate the observation that there tends to be an inverse relationship between VOT and duration of the following vowel (Summerfield, 1981). However, to approximate this inverse relationship in natural production, the extent of vowel shortening was not entirely commensurate with changes in VOT, that is, for every 1 ms of VOT increase, the vowel was shortened by less than 1 ms (Allen and Miller, 1999; Toscano and McMurray, 2010). The default vowel-VOT ratio of .65 from Winn (2020) was used for modeling this trade-off relation.

---

<sup>5</sup>[https://github.com/ListenLab/VOT/blob/master/Make\\_VOT\\_Continuum\\_v31.txt](https://github.com/ListenLab/VOT/blob/master/Make_VOT_Continuum_v31.txt)

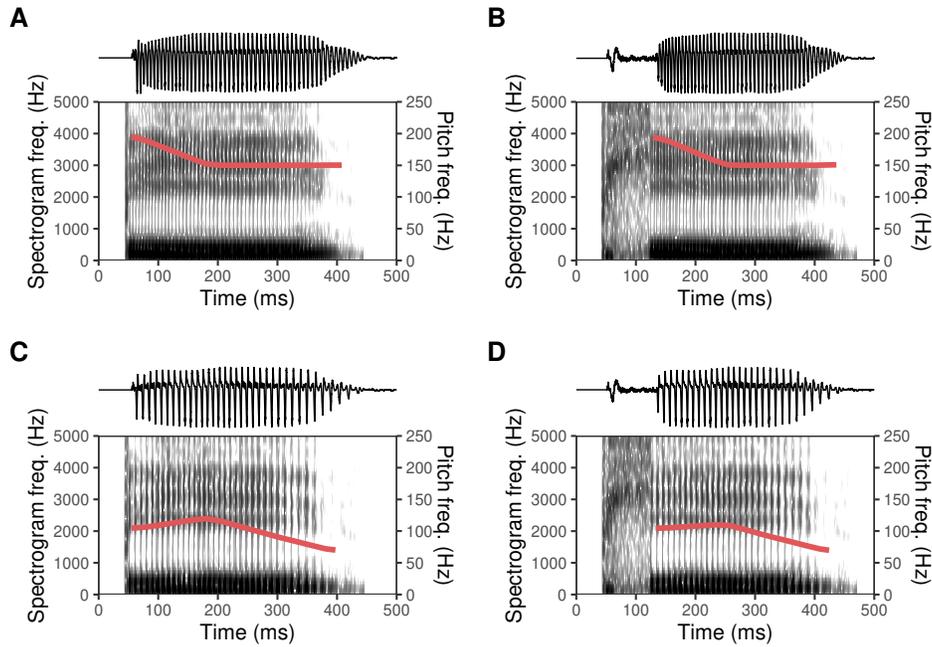
*F0*: *F0* was set at one of the seven values, from 105 Hz to 195 Hz with a step size of 15 Hz, at the beginning of the vowel. *F0* then rose/fell linearly for the following 122 ms (or 35% of the vowel duration) to 150 Hz for Tone 1 stimuli and to about 118 Hz for Tone 4 stimuli. The step size was set to 15 Hz so that the difference in *F0* would be large enough to be noticeable but not too large so as to distort the *F0* trajectory significantly, and the temporal extent of manipulation was fixed at 35%, following the practice in Guo (2020), which was in turn based on the Mandarin production data in her study.

The waveform and spectrograms of two example target syllables are shown in Figure 3.10.



**Figure 3.9:** Manipulation of target stimuli for all perception experiments. **A.** Each dot represents one stimulus, with its  $x$  coordinate corresponding to the voice onset time (VOT) of the initial labial stop, and its  $y$  coordinate to the initial *F0* of the following vowel. **B.** Illustration of *F0* trajectory manipulation for target syllables. Note the vowel duration in actual stimuli is not necessarily 350 ms due to the trade-off between VOT and vowel duration, which was also manipulated. The invariant parts across different tokens are shifted vertically in the figure for visual clarity only.

The creation of filler items roughly followed the same first two steps in creating the target items (e.g., [i<sup>1</sup>] and [i<sup>4</sup>]) were created from a natural production of *yi1*, except that the filler [mi<sup>4</sup>] was modified from a natural Tone 4 syllable, *mi4*, rather than being constructed from the Tone 1 syllable *mi1*. However, the tonal contour



**Figure 3.10:** Sample target stimuli. **A.** Tone 1, with VOT of 0 ms and F0 of 195 Hz. **B.** Tone 1, with VOT of 80 ms and F0 of 195 Hz. **C.** Tone 4, with VOT of 0 ms and F0 of 105 Hz. **D.** Tone 4, with VOT of 80 ms and F0 of 105 Hz. The red line in each figure represents the F0 trajectory.

of this filler item was similarly styled to that of target Tone 4 items, to prevent this filler from standing out from the other stimuli. The rationale behind was to add acoustic variability to stimuli and therefore to encourage the participant to abstract away from low-level acoustic signals. Note, however, that this decision is not critical with regard to data analysis, as only data from target stimuli were included.

### English Stimuli

The target stimuli for the English version of the perception experiment were identical to those for the Mandarin version. The filler stimuli, on the other hand, were changed to [mi̯], [mi̯\] (similar to English *me*), [ni̯], and [ni̯\] (similar to English *knee*). The reason why [i̯] and [i̯\] were not used was to avoid the use of letter *E*

as one of the response options; it was preferable that all four response options were lexical items.

### 3.4.3 Procedure

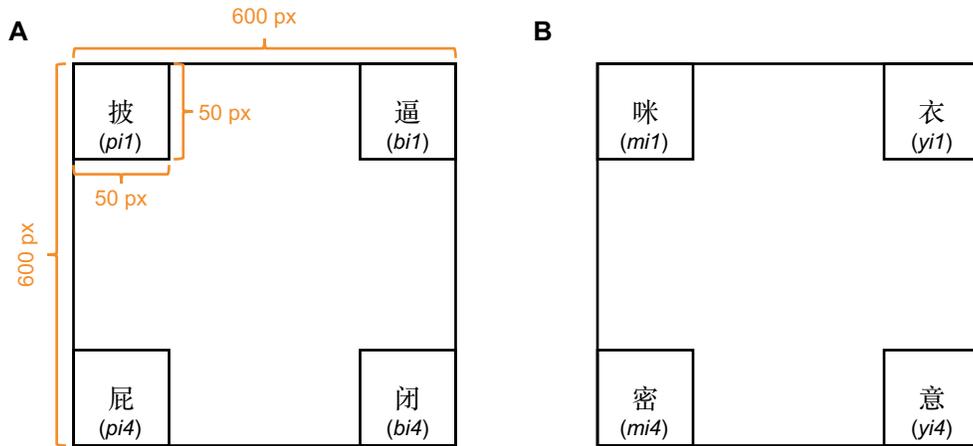
In presenting the experimental procedure, I first go through the configuration and layout of response options in each trial, and then described the task involved. At a high level, the task was a forced-choice identification task, where the participant clicked on one word out of a choice of four.

#### Mandarin Trial Configuration

Experimental trials consisted of two trial types: targets and fillers, depending on whether the audio stimulus being played were from the target or filler set. Both trial types had as response options four Mandarin monosyllabic words. For the targets, the four response words were *pi1* 披, *pi4* 屁, *bi1* 逼, and *bi4* 闭, which differed from one another in stop voicing and lexical tone. Note that these words were also included in the production stimuli. The four options were placed at the four corners of a 600 px × 600 px square, with each option having a response area of a 50 px × 50 px square, as illustrated in Figure 3.11. Furthermore, the relative positions of the four options were constrained in such a way that two words distinguished only in the voicing of onset (e.g., *pi1* vs. *bi1*) were always next to each other, so there were only 16 (4 sides × 4 possible positionings/side) possible trial option configurations. The 16 trial configurations were counterbalanced across participants at the time of testing (i.e., the counterbalance was not taken into account when participants' data was selected for analyses), and the same configuration was used throughout the course of experiment. The decision to maintain the same configuration was to prevent the participant from doing visual search, which might introduce additional cognitive load.

For the fillers, the four options were *yi1* 衣, *yi4* 意, *mi1* 咪, and *mi4* 密, which similarly differed in both onset and lexical tone. However, their positioning was not constrained in any manner, as the data collected in filler trials were not analyzed. This resulted in 24 (= 4!) possible configurations, and each participant was randomly assigned a configuration, which remained the same throughout the entire

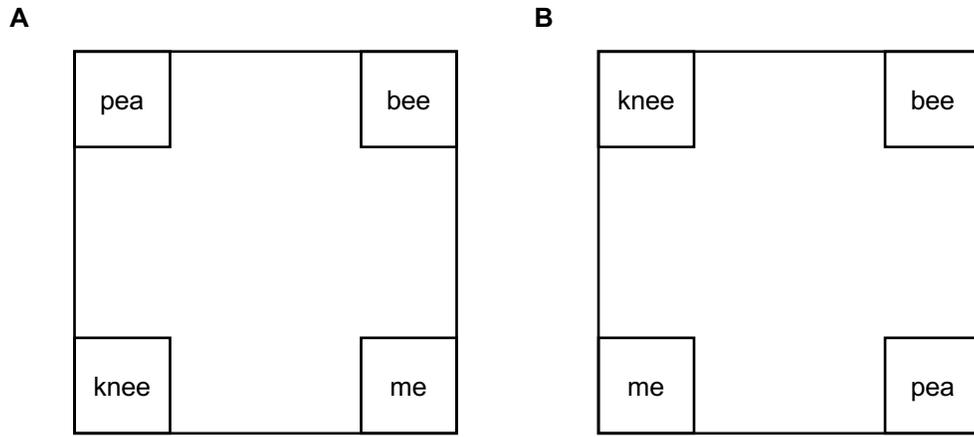
experiment.



**Figure 3.11:** One of possible response configurations for the Mandarin perception experiment. **A.** Target trial. **B.** Filler trial.

### English Trial Configuration

The experimental trials for English similarly consisted of target trials and filler trials. However, unlike the Mandarin version, the two trial types differed from each other only in the audio stimulus being played; that is, the same response layout was used for both trial types. This being the case came from the fact that English lacks lexical tone, so it was impossible to have a response layout parallel to that in the Mandarin version. The trial configuration always had as response options four English words: *pea*, *bee*, *me*, and *knee*. The four words were arranged such that *pea* and *bee* were always only one edge away from each other (and as a consequence *me* and *knee* were likewise always next to each other)—the same constraint that phonological competitors in terms of stop voicing were always adjacent to each other. This resulted in 16 possible option configurations (4 sides  $\times$  4 arrangements/side), two of which are shown in Figure 3.12. These 16 configurations were counterbalanced across participants at the time of testing, and the configuration remained unaltered within an experiment session.



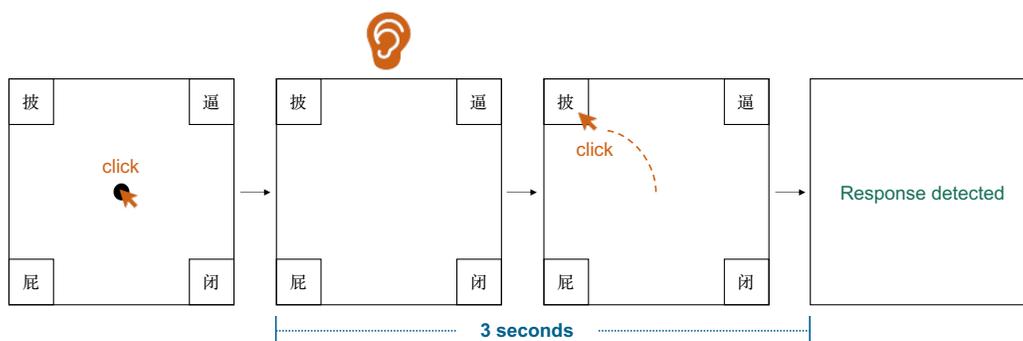
**Figure 3.12:** Two possible response configurations for the English perception experiment. **A.** With the responses corresponding to target audio stimuli on top. **B.** With the responses corresponding to target audio stimuli on left.

### Task Procedure

The experiment procedure was the same for both the Mandarin and English versions of the experiment. The whole experiment took place online and was programmed in jsPsych (de Leeuw, 2015). Participants were encouraged to use a physical mouse and to wear headphones for the experiment, though they could also do the experiment with a touchpad and/or the built-in loud speakers on their computer. The experiment started with a short hearing test, where the participant had to select the quietest tone out of three tones differing in loudness (Woods et al., 2017). This test was challenging to do when *not* wearing headphones. They had to respond correctly in at least five out of six trials to pass the test.

The basic procedure followed that of Experiment 1 from Dale et al. (2007). During each trial, the four options were first presented for 500 ms to remind the participant of the word at each corner. Next, a black dot, the radius of which was 5 px, appeared in the center of the screen, which the participant had to click for the audio stimulus to be immediately presented. The function of this center dot was to ensure that the mouse cursor was reset to (approximately) the center. The participant then had a three-second period to indicate their response by clicking

one of the words.



**Figure 3.13:** Illustration of trial procedure in the Mandarin perception experiment. The procedure follows a typical static starting procedure in mouse-tracking experiments, where an audio stimulus is played upon the participant’s clicking on the center dot, and they have to click on an option within three seconds counting from the onset of audio stimulus presentation.

Participants had to go through three blocks, with each block having the same tokens and differing only in the order in which the tokens were presented. To have a target-to-filler ratio of about 4:1, each block contained one repetition of target stimuli and seven repetitions of filler stimuli, resulting in a total of 126 ( $= 98 \times 1 + 7 \times 4$ ) trials in each block. Three blocks were used to achieve a compromise between having as many trials as possible and limiting the duration of the experiment under 30 minutes. Between blocks the participant could take a self-timed break.

#### 3.4.4 Additional Participant Inclusion Criteria

As mentioned in Section 3.3.1, participants’ performances in the perception experiment formed a part of the inclusion criteria. The purpose is to only include participants who actually paid attention during the experiment. This criterion was operationalized by first calculating by-participant “correct” percentage of responses for each language version, separated for target and filler trials. For the target trials in the Mandarin perception experiment, a correct trial was a target trial where the participant selected as the response a word whose tone matched the tonal contour of the audio stimulus. For the filler trials in the Mandarin experiment, a correct trial

was a filler trial whose selected response word corresponded exactly to the audio stimulus (e.g., selecting *yil* for [iɿ]). For the target trials in the English version of the experiment, a correct trial was a target trial whose response was either *pea* or *bee*. For the filler trials in the English experiment, a correct trial was defined as a filler trial which had *me* or *knee* as the response, taking into account the fact that the bilabial and alveolar nasal onsets in the filler stimuli were perceptually confusable. For a participant who completed both English and Mandarin perception experiments, four percentage scores were computed—% correct for targets in Mandarin perception, % correct for fillers in Mandarin perception, % correct for targets in English perception, and % correct for fillers in English perception. For each participant, an average correct percentage across the four language/trial type combinations was computed. Participants were then ranked based on the average correct percentage in a descending order, and the data from the top 25 participants was included in the analyses.

### 3.4.5 Omitted Data

For both Mandarin and English versions of the perception experiment, only the response data from the target trials were considered. Additionally, only the “correct” target trials, as defined in Section 3.4.4 above, were included in the analyses. Altogether, 216 (129 Tone 1 tokens and 87 Tone 4 tokens) out of 7,350 target trials were removed from the Mandarin experiment, and 59 (29 Tone 1 tokens and 21 Tone 4 tokens) out of 7,350 target trials were removed from the English experiment.

### 3.4.6 Statistical Analyses

A variant of logistic regression was used to derive the perceptual weight for post-stop F0. Note that, to use logistic regression, the choice the participant was making (i.e., four options varying along lexical tones and stop voicing) was mapped down to a binary *voicing* identification. This mapping assumes that the lexical tone is clear from the F0 contour of the audio stimulus, so participants’ responses can be approximated as a binary p<sup>h</sup>/p choice. In all the models, participants’ responses were modeled as a function of VOT, post-stop F0, and tonal categories. The coefficient of the post-stop F0 variable was then used as its perceptual weight. Similar to

the production models, all models were fitted with Bayesian mixed-effects models using `CmdStanR` (Gabry and Češnovar, 2021).

### Variables

Before being fed into the analyses, the two continuous predictor variables—**VOT** and **post-stop F0**—were  $z$ -transformed with respect to the original sequence (e.g., the VOT value of 0 was consistently mapped to  $[0 - \text{Mean}(0, 13, 27, 40, 53, 67, 80)]/\text{SD}(0, 13, 27, 40, 53, 67, 80) = -1.39$ , regardless of listener).  $Z$ -transformation was used so that the scale is parallel to that in the production experiment. The variable **tone** was sum-coded with TONE 1 and TONE 4 being coded with 1 and  $-1$  respectively. The variable **tone** was included because the responses were mapped from four alternatives to a binary choice (see above). Also, since post-stop F0 and tone were manipulated independently, one cannot causally influence the other. Both variables were therefore included. The default level for the response was always /b/ (i.e., the /b/ response was coded with 0, and the /p/ response was coded with 1), so a positive coefficient for a given predictor variable means that higher values of this dimension elicit more voiceless responses in listeners than lower values.

### Model Structure

Listeners' responses were assumed to be generated by a mixture of two different sources: one source was the logistic function of terms formed with the predictors, and the other was sheer randomness or guessing due to the listener not paying attention or accidentally making a mistake, that is, the response came from one of the four options being selected by chance (Kruschke, 2015). Formally, each response had a chance,  $\gamma$ , of being generated by the guessing process, and, with probability  $1 - \gamma$ , the response came from the logistic function of the predictor:

$$\text{/p/ response} \sim \text{bernoulli} \left( \gamma \cdot \frac{1}{4} + (1 - \gamma) \cdot \text{logistic} \left( \beta_0 + \sum_i \beta_i x_i \right) \right).$$

Model fitting thus involved estimating the guessing probability  $\gamma$  along with the

logistic parameters,  $\beta_i$ , which were taken to represent the weight given to each dimension in categorization. Bayesian hierarchical models were employed to derive a posterior probability distribution for each parameter. The full model consisted of two submodels with the same parameterization and predictors: one submodel predicted listeners' responses in the Mandarin mode while the other submodel predicted listeners' responses in the English mode, and the two submodels were tied together by correlating all logistic parameters with one another in a multinormal distribution. A guessing probability was estimated for each listener in each language mode independently. Logistic parameters were parameterized such that each was decomposed into a fixed-effect part, corresponding to the weight at the population level, and a random-effect part, representing the adjustment for each listener.

Each model used 4,000 samples across four Markov chains and was fit with a regularizing prior of  $\text{Normal}(\mu = 0, \sigma = 10)$  for the fixed-effect estimates. An  $\text{Exponential}(r = 1)$  distribution was used as the prior for listener-specific adjustments. Correlations among listener-specific adjustments used the LKJ prior with  $\xi = 1$ . The guessing probability for each listener in each language had a uniform prior between 0 and 1. All models showed no divergent transitions, and sampling chains were well-mixed (i.e., all  $\hat{R} < 1.01$ ). The detailed mathematical specifications for the final model can be found in Section D.2.2 in the appendix.

### **Candidate Models**

Similar to the statistical models for production data, candidate models for perceptual performance reflected both prior knowledge and a compromise between complexity and predictive accuracy. Given that VOT is the primary cue for the stop voicing contrast in Mandarin and English, all the models in the comparison had VOT automatically included, with the simplest model containing VOT as the sole predictor. Built off this simplest models were candidates with increasing complexity introduced by terms involving post-stop F0 and tone. The full list of models considered is listed in Table 3.9.

**Table 3.9:** Candidate perceptual models considered in model comparison, with their ELPD-LOO means and standard deviations.

Model	ELPD-LOO	ELPD-LOO standard error	Predictors
M1	-1419.5	52.3	VOT
M2	-1366.2	51.3	VOT + F0
M3	-1395.4	52.0	VOT + tone
M4 (final)	-1340.5	51.4	VOT + F0 + tone
M5	-1334.6	51.5	VOT + F0 + tone + F0 × VOT
M6	-1325.4	51.7	VOT + F0 + tone + F0 × tone
M7	-1326.0	51.8	VOT + F0 + tone + F0 × VOT + F0 × tone
M8	-1327.9	52.1	VOT + F0 + tone + F0 × VOT + F0 × tone + VOT × tone

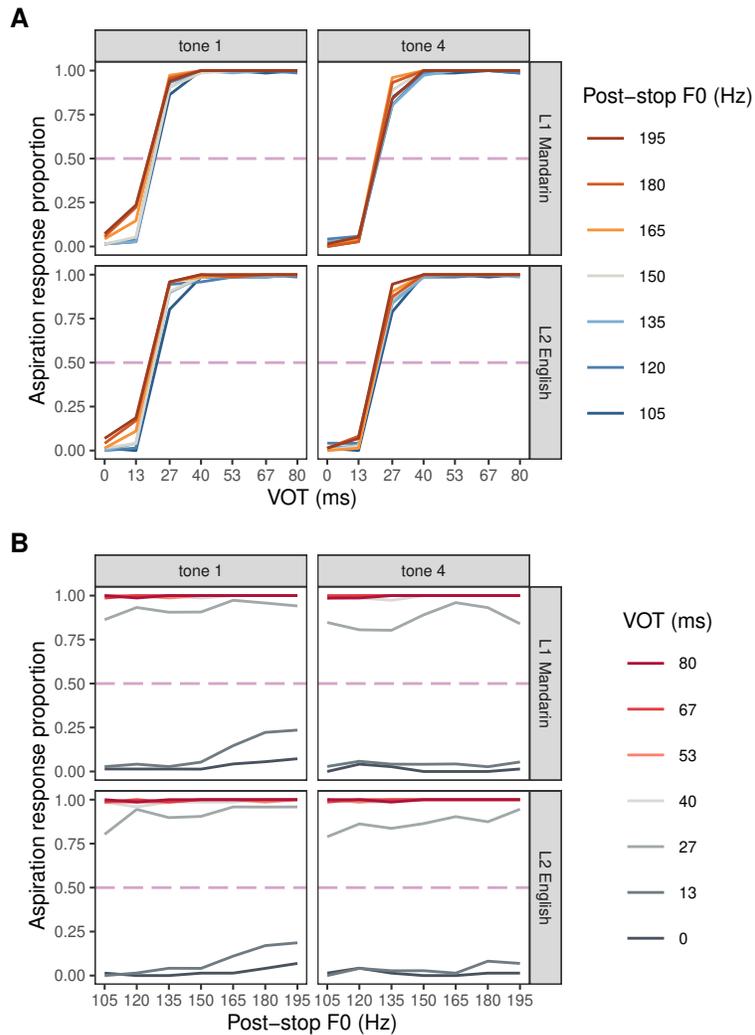
### 3.4.7 Results: Perceptual Weights of Post-Stop F0

The response patterns across different VOTs, post-stop F0s, tones, and experiment versions are shown in Figure 3.14. The ELPD-LOO mean and standard error for each candidate model are listed in Table 3.9, and the model comparison results among the candidate models are detailed in Table 3.10.

Model comparison indicated the importance of post-stop F0 and tone in predicting listeners' categorization performances (M1 vs. M2 and M3 vs. M4 for post-stop F0; M1 vs. M3 and M2 vs. M4 for tone). However, including interaction terms between any pairs of the cues did not lead to substantial increase in predictive accuracy. For this reason, M4 was selected as the final model, and subsequent discussion was made on the basis of M4.

### Population Results

The marginal posterior distributions for population-level effects from M4 are summarized in Table 3.11. All predictors, including the intercepts, had an effect on categorization. The cue of most interest here is post-stop F0, but for completeness, the results for other dimensions are also briefly discussed. On the basis of the fact that the 89% CrIs for post-stop F0 did not contain 0 in both Mandarin and English (Mandarin:  $\bar{\beta} = .53$ , 89% CrI = [.30, .75],  $p(\beta > 0) = 1.00$ ; English:  $\bar{\beta} = .89$ , 89% CrI = [.64, 1.14],  $p(\beta > 0) = 1.00$ ), post-stop F0 was judged to



**Figure 3.14:** Line charts of Mandarin-English bilinguals' aggregated categorization of word-initial stops in each language, shown as a function of VOT, post-stop F0, and tone. **A.** With VOT on the *x*-axis. **B.** With post-stop F0 on the *x*-axis to highlight its effect on categorization.

**Table 3.10:** Model comparison results for key perception model pairs, expressed in differences in ELPD-LOO and associated standard errors. Pairs judged to differ in predictive power are marked by asterisks.

Model	M2	M3	M4	M5	M6	M7	M8
M1	-53.2* (10.8)	-24.1* (7.5)					
M2			-25.8* (8.1)				
M3			-54.9* (11.2)				
M4				-5.8 (6.6)	-15.1 (7.7)		
M5						-8.6 (6.6)	
M6						.6 (4.8)	
M7							1.8 (2.7)

be a cue for stop voicing in both languages. However, the weight assigned to this cue was language-dependent, as evidenced by the 89% CrI of difference in post-stop F0 weights occupying only negative values ( $\bar{\beta} = -.36$ , 89% CrI =  $[-.67, -.04]$ ,  $p(\beta < 0) = .97$ ). In particular, listeners relied on post-stop F0 more when the stimuli were presented as English words than when the exactly same stimuli were perceived as Mandarin words. The magnitude of the intercept was indicative of the location of the category boundary: a positive intercept meant there were more /p/ responses in general, which translated to an early boundary within the range of values considered. This can be clearly seen in Figure 3.14, where the category boundary in terms of VOT (i.e., the VOT value where the proportion of /p/ responses is .5) occurs before the midpoint of the VOT continuum. Also, the intercept seemed stable across participants' Mandarin and English categorization performances. VOT, as expected, was the strongest cue for the voicing decision, and its weight was comparable across languages. Finally, Tone 1 stimuli seemed to trigger more voiceless responses to a similar degree in both languages.

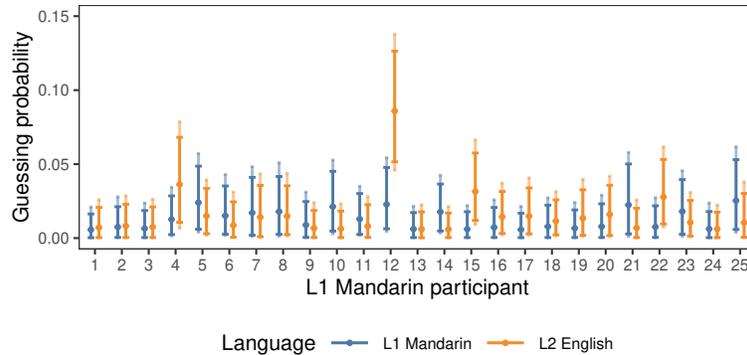
**Table 3.11:** Marginal posterior summary for key population-level parameters from M4. The parameters whose effects are judged to be strong are marked with \*\*, and those whose effects are judged to be weak are marked with \*.

Parameter	Mean	SD	89% CrI	$p(\text{dir.})$
intercept <sub>Man</sub> **	9.14	.68	[8.11, 10.28]	$p(\beta > 0) = 1.00$
VOT <sub>Man</sub> **	13.81	1.07	[12.19, 15.63]	$p(\beta > 0) = 1.00$
F0 <sub>Man</sub> **	.53	.14	[.30, .75]	$p(\beta > 0) = 1.00$
tone <sub>Man</sub> **	.54	.12	[.35, .73]	$p(\beta > 0) = 1.00$
intercept <sub>Eng</sub> **	9.88	.73	[8.81, 11.07]	$p(\beta > 0) = 1.00$
VOT <sub>Eng</sub> **	15.08	1.11	[13.42, 16.9]	$p(\beta > 0) = 1.00$
F0 <sub>Eng</sub> **	.89	.16	[.64, 1.14]	$p(\beta > 0) = 1.00$
tone <sub>Eng</sub> **	.42	.12	[.22, .61]	$p(\beta > 0) = 1.00$
intercept <sub>Man</sub> – intercept <sub>Eng</sub>	–.74	.92	[–2.16, .74]	$p(\beta < 0) = .78$
VOT <sub>Man</sub> – VOT <sub>Eng</sub>	–1.28	1.41	[–3.43, 1.08]	$p(\beta < 0) = .81$
F0 <sub>Man</sub> – F0 <sub>Eng</sub> **	–.36	.20	[–.67, –.04]	$p(\beta < 0) = .97$
tone <sub>Man</sub> – tone <sub>Eng</sub>	.12	.17	[–.15, .39]	$p(\beta > 0) = .75$
$\rho_{\text{intercept}_{\text{Man}}, \text{intercept}_{\text{Eng}}}$ **	.41	.21	[.04, .74]	$p(\beta > 0) = .96$
$\rho_{\text{VOT}_{\text{Man}}, \text{VOT}_{\text{Eng}}}$ **	.52	.19	[.20, .79]	$p(\beta > 0) = .99$
$\rho_{\text{F0}_{\text{Man}}, \text{F0}_{\text{Eng}}}$	.34	.28	[–.15, .73]	$p(\beta > 0) = .88$
$\rho_{\text{tone}_{\text{Man}}, \text{tone}_{\text{Eng}}}$	.10	.33	[–.44, .62]	$p(\beta > 0) = .62$
$\rho_{\text{VOT}_{\text{Man}}, \text{F0}_{\text{Man}}}$ *	–.33	.24	[–.70, .08]	$p(\beta < 0) = .90$
$\rho_{\text{VOT}_{\text{Eng}}, \text{F0}_{\text{Eng}}}$	–.06	.27	[–.50, .38]	$p(\beta < 0) = .59$
$\rho_{\text{tone}_{\text{Man}}, \text{F0}_{\text{Man}}}$	.20	.31	[–.32, .66]	$p(\beta > 0) = .75$
$\rho_{\text{tone}_{\text{Eng}}, \text{F0}_{\text{Eng}}}$	.11	.30	[–.40, .59]	$p(\beta > 0) = .65$
$\rho_{\text{tone}_{\text{Man}} - \text{tone}_{\text{Eng}}, \text{F0}_{\text{Man}} - \text{F0}_{\text{Eng}}}$	–.11	.34	[–.67, .42]	$p(\beta < 0) = .63$

### Individual Results

The guessing probability estimated for each listener in each language is plotted in Figure 3.15, and the detailed numbers are listed in Table D.5 in the appendix. Overall, the guessing probabilities were very low, with 24 out of 25 listeners having a mean guessing probability below 5% in either language and only one listener (i.e., MS12) having a value of around 10% for the English task.

Individual listeners' weights for various cues and the weight differences in



**Figure 3.15:** Individual participants’ guessing probabilities in each language. The dots represent the posterior means while the error bars stand for the 89% CrI.

these cues across languages are visualized in Figure 3.16. Table D.6 and Table D.7 in the appendix provide the numerics the figure is based on. Again, the results regarding the cue weight for post-stop F0 are discussed first, as it is the dimension of interest here; the results on other cues are also summarized in passing for completeness.

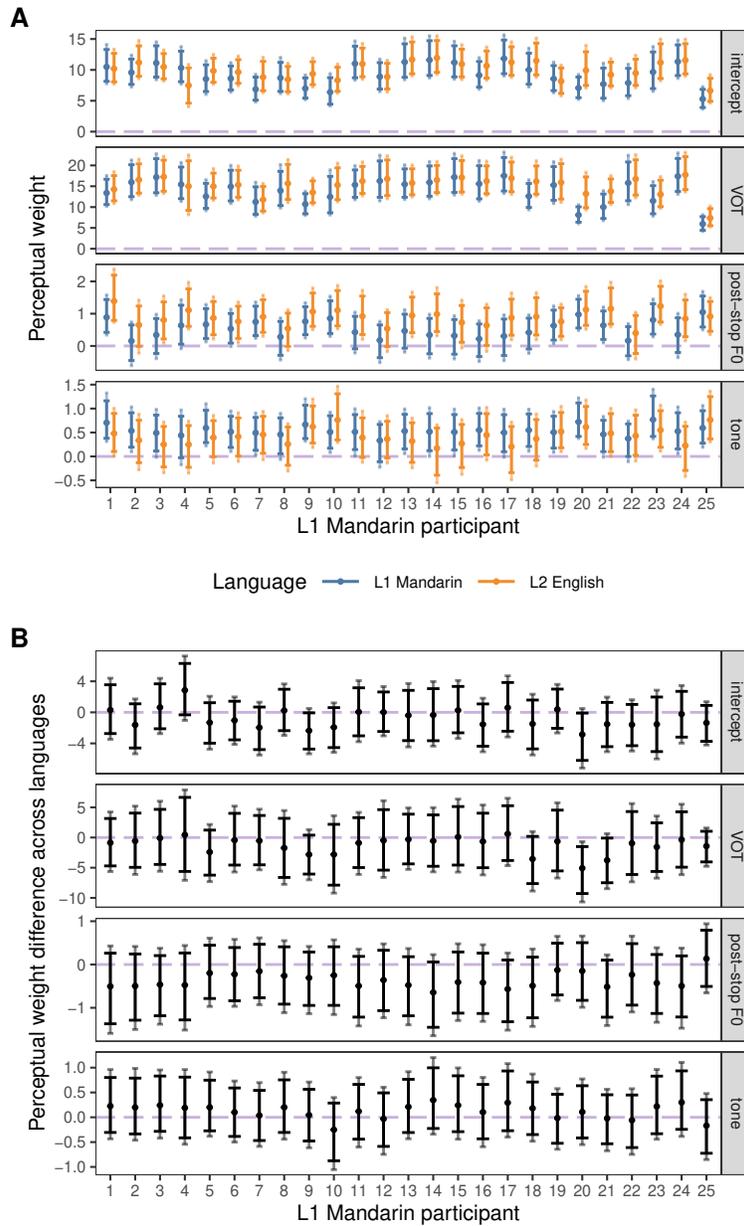
As shown in the [post-stop F0] panel of Figure 3.16A, though the 89% CrI for the post-stop F0 weight did cross 0 for some listeners, all listeners had a positive mean weight for the post-stop F0 cue for both languages, signifying that, generally speaking, the chance the /p/ response was selected went up with an increasing post-stop F0. Comparing the weights of this cue across languages (Figure 3.16B), all but one listener (i.e., participant 25) had a higher mean weight in English than in Mandarin; however, because of the relatively large uncertainty surrounding the estimated weight values, there is no strong evidence that there is a difference on an individual level, as signified by the 89% CrI for the *difference* between the weights containing 0 for all participants. In spite of this large uncertainty, the trend seemed robust and echoed the population-level pattern in terms of the direction of the effect. Another way to understand the cue is to examine whether the cue use is consistent across languages at the individual level by correlating the weights from the two language contexts. In fact, the correlation information can be directly read off from the fitted model and is summarized in the last few row in Table 3.11

and visualized in Figure 3.17. As can be seen in Figure 3.17C, there was a weak positive correlation of this cue across languages ( $\bar{\rho} = .34$ , 89% CrI =  $[-.15, .73]$ ,  $p(\rho > 0) = .88$ ), though the 89% CrI for this correlation also spilled to the negative side, probably due to the small number of participants, which was not effective in constraining the uncertainty when the correlation was weak. There also seemed to be a trading relationship between the VOT and post-stop F0 weights in Mandarin (Figure 3.17E;  $\bar{\rho} = -.33$ , 89% CrI =  $[-.70, .08]$ ,  $p(\rho < 0) = .90$ ), but not in English (Figure 3.17F;  $\bar{\rho} = -.06$ , 89% CrI =  $[-.50, .38]$ ,  $p(\rho < 0) = .59$ ).

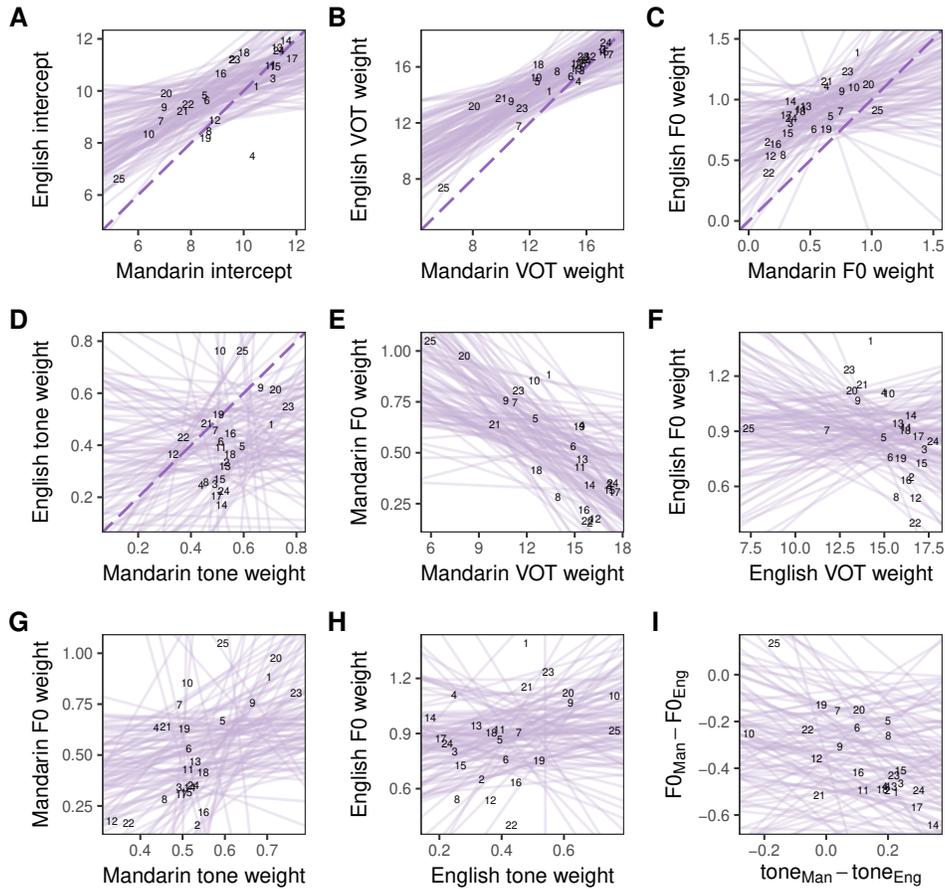
For the intercepts, which were connected with the location of category boundary, even though individual listeners varied with respect to the boundary location, the location was relatively stable within a listener, as evidenced from Figure 3.17A and from the positive 89% CrI of the correlation coefficient ( $\bar{\rho} = .41$ , 89% CrI =  $[.04, .74]$ ,  $p(\rho > 0) = .96$ ). The same story could be stated for the VOT cue: individuals varied in a structured way, with the cue use being stable within the same individual across contexts ( $\bar{\rho} = .52$ , 89% CrI =  $[.20, .79]$ ,  $p(\rho > 0) = .99$ ). As for tone, it seemed that, for most listeners (19 out of 25), the effect of Tone 1 stimuli eliciting more voiceless responses was stronger in Mandarin than in English, though the difference was not particularly big.

### 3.4.8 Comparing Individual Post-Stop F0 Weights across Production and Perception

It might be expected that the weight of a given acoustic dimension on a speaker's production would predict the weight assigned to that dimension in the same speaker's perception. To test this hypothesis empirically, individuals' post-stop F0 weights in production, which were represented by each person's mean Cohen's  $d$ , were correlated with their post-stop F0 weights in perception, as approximated by the beta-coefficient for F0 in the logistic regression model, separately for each language. Figure 3.18 shows the perceptual weights plotted against the production weights. The results from the correlation analyses were dependent on the language, with no correlation across production and perception for Mandarin ( $\bar{\rho} = -.02$ , 89% CrI =  $[-.34, .31]$ ,  $p(\rho < 0) = .54$ ) but a positive correlation for English ( $\bar{\rho} = .39$ , 89% CrI =  $[.09, .64]$ ,  $p(\rho > 0) = .98$ ). However, it should be cautioned the correlation was done on the posterior *means* of the estimated weights, disregarding

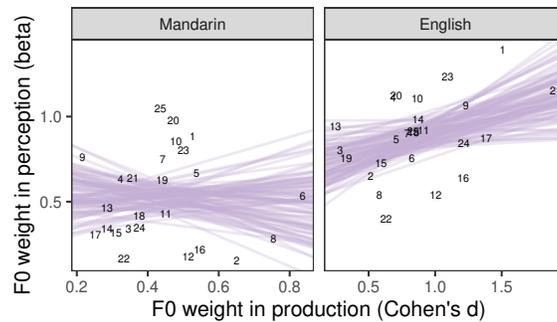


**Figure 3.16:** Individuals' estimated weights from the perceptual model. **A.** Distributions of individual weights along various dimensions for Mandarin and English. **B.** Differences in cue weights along the same dimension across languages. In both figures, posterior means are represented by the dots. The 89% CrIs are marked by the inner error bars, while the 95% CrIs are marked by the outer error bars.



**Figure 3.17:** Scatter plots showing relationships (or lack thereof) between various cues. **A.** Category boundaries across languages. **B.** VOT weights across languages. **C.** Post-stop F0 weights across languages. **D.** Tone weights across languages. **E.** F0 vs. VOT in Mandarin. **F.** F0 vs. VOT in English. **G.** F0 vs. tone in Mandarin. **H.** F0 vs. tone in English. **I.** Differences in post-stop F0 weights vs. differences in tone weights. Solid lines represent 100 regression lines fit with 100 posterior draws, to show direction and uncertainty in the correlation. The dashed line in **A-D** is  $y = x$ , where the intercept or VOT / post-stop F0 / tone weight for Mandarin equals that for English.

any uncertainty surrounding these estimates, so the apparent correlation (or lack of correlation) might not be reliable. To properly account for the uncertainty requires fitting both production and perception data with a single model, which demands more sophisticated statistical models and therefore is left for future research.



**Figure 3.18:** Post-stop F0 weightings across perception and production for each participant in Mandarin and English. Solid lines represent 100 regression lines fit with 100 posterior draws, to show direction and uncertainty in the correlation.

## 3.5 Discussion

### 3.5.1 Summary of Results

The current study explores the ambiguity of F0 in Mandarin through L1 Mandarin-L2 English bilinguals' production and perception of the stop voicing contrast in their L1 and L2. The results from the conducted experiments are summarized in Table 3.12, which ties them back to the hypotheses and predicted results listed in Table 3.2, and discussed below. At the population level, these results largely echoed recent work by Guo (2020).

In both their Mandarin and English productions, the post-stop F0 following an aspirated stop tended to be higher than that following an unaspirated stop, and unaspirated stops in turn induced a higher F0 than sonorants. In addition, the extent to which post-stop F0 was differentiated between the aspirated and unaspirated categories hinged on the language and lexical tone: comparing English with Mandarin

(which was represented as an average between Tone 1 and Tone 4 in this study), English supported a bigger post-stop F0 difference; contrasting Tone 1 and Tone 4 in Mandarin, Tone 4, which was realized with a higher F0 register phonetically, also sustained a slightly greater post-stop F0 distinction. The production weights for post-stop F0 across languages was also reflective of the finding above: post-stop F0 assumed a larger weight in English than in Mandarin, both at the population level and for most individuals (19 out of 25 speakers). These findings therefore support the view that post-stop F0 perturbation is not necessarily intrinsic to the articulatory system.

In perception, post-stop F0 was also used as a cue for stop voicing by the same L1 Mandarin-L2 English participants when put in either a Mandarin or an English context. However, the language context modulated the weight such that post-stop F0 carried more weight when the stimuli were presented as English words than when the same stimuli were presented as Mandarin words. This language-conditioned change in cue weighting was statistically well-supported at the population level, but, at the individual level, because of fewer data points (i.e., the same stimuli were only repeated three times for each participant), the model was less confident. Nonetheless, almost all individuals (24 out of 25) followed the population trend as far as posterior means were concerned. Overall, the patterns revealed in the perception experiment are supportive of the claim that L2 learners can adjust the use of a cue in different language contexts.

Compared across production and perception, on a population level, a higher production weight for post-stop F0 mapped to a higher perceptual weight for the same cue. This is reflected in the bilinguals' relying more on post-stop F0 to contrast stop voicing in English than in Mandarin across modalities. On an individual level, on the other hand, an individual's production weight did not reliably predict the same individual's perceptual weight, at least for post-stop F0 with the adopted metrics in Mandarin. This mismatch therefore suggests at least some independence of the two modalities.

**Table 3.12:** Predicted production and perception results under difference hypotheses.

<b>Production</b>		
<b>Hypotheses</b>	<b>Predicted production results</b>	<b>Match actual results?</b>
Post-stop F0 purely due to physiological / aerodynamic reasons (e.g., Kohler, 1984; Ladefoged, 1967; Ohala and Ohala, 1972) or total transfer of post-stop F0 cue use in Mandarin to English, as predicted by the SLM-r	Post-stop F0 difference the same in Mandarin and English tokens	✗
Post-stop F0 partially subject to active controlling (Kingston and Diehl, 1994)	The extent of post-stop F0 difference might depend on the language (i.e., larger in English than in Mandarin)	✓ Post-stop F0 difference between aspirated and unaspirated stops was bigger in English than in Mandarin at the population level and for many speakers.
<b>Perception</b>		
<b>Hypotheses</b>	<b>Predicted perception results</b>	<b>Match actual results?</b>
Transfer of the Mandarin cue-weighting strategy to English, as predicted by the SLM-r and PAM-L2	Post-stop F0 weights the same across Mandarin and English	✗
Flexibility in cue use: attributing variation in post-stop F0 partially to lexical tone and partially to stop voicing in Mandarin, but only to stop voicing in English	Post-stop F0 weights depend on the language context (i.e., a higher weight in English than in Mandarin)	✓ Post-stop F0 carried more weight in English than in Mandarin at the population level. The model was less confident at the individual level, though the trend was the same as the population result for 24 out of 25 listeners.

### 3.5.2 Flexibility of Cue-Weighting in L2

The findings from the experiments show that bilinguals, even non-early/non-simultaneous/non-child bilinguals, are able to dynamically adjust their cue-weighting strategies in facing different language contexts in production as well as perception. Prior demonstrations on bilinguals' ability to fine-tune the use of various acoustic dimensions concerned mainly simultaneous or early bilinguals (e.g., Antoniou et al., 2010, 2012; Gonzales et al., 2019; Gonzales and Lotto, 2013). However, as reviewed in Section 3.2.5, more recent works have suggested that late L2 learners are also capable of such a deed (Amengual, 2021; Casillas and Simonet, 2018). The results from this study are in line with these recent works in that Mandarin-English bilinguals shift the post-stop F0 weight in response to the current language mode. Crucially, however, this study also demonstrates bilinguals' capability to modulate the use of a secondary cue, as opposite to just the primary cue as in previous works.

### 3.5.3 Role of Tone in Post-Stop F0

The fact that, in production, greater post-stop F0 difference was found in Tone 4, which was realized with a higher initial pitch than Tone 1, and that, in perception, Tone 1 syllables induced more aspirated responses, points to a potential role of tone identity in conditioning post-stop F0. In fact, previous works have documented such cases in production at least. For example, as mentioned in Section 3.2.2, Guo (2020) reports that F0 following an aspirated stop is higher only in Tone 1 and Tone 4 syllables (both of which begin with a high pitch register) while F0 following an *unaspirated* stop is higher in Tone 2 and Tone 3 syllables (both having a low initial register). Kirby (2018) investigates the post-stop F0 effects in two other tonal languages—Thai and Vietnamese—and finds that the greatest post-stop F0 effects for Thai are present in the high-falling tone environment, though the results from Vietnamese are less clear-cut. Even in non-tonal languages, post-stop F0 difference is most prominent in high-pitch, focused conditions (Hanson, 2009; Kirby and Ladd, 2016). The enlargement of post-stop F0 difference in high-pitch contexts across tonal and non-tonal languages suggests that a general, language-independent explanation in terms of F0 control might be responsible, and more

research is needed to elucidate this hypothesis.

With respect to perception, a careful inspection of Figure 3.14 reveals that increased /p/ responses in Tone 1 tokens resulted largely from higher post-stop F0 values in Tone 1 provoking more /p/ responses when VOT was ambiguous (i.e., when VOT was around 13 ms). A possible explanation for why Tone 1, as compared with Tone 4, led to such an effect is that it is not just the initial value of F0 that matters; the listener also tracks changes in F0 slope throughout the syllable, and such changes also contribute to the perception of aspiration. In the context of the current perception experiment, all Tone 1 tokens end with a tailing flat F0 contour, which might enhance the percept of the initial drop in F0, whereas the falling F0 contour in Tone 4 tokens might perceptually offset the initial drop in F0, resulting in the change in F0 being less noticeable. Another explanation is that since Tone 4 syllables tend to have a higher initial F0 in production than Tone 1 syllables, Mandarin listeners might require an acoustically higher initial F0 value in Tone 4 tokens to judge a token as starting with a high F0. Of course these speculations await experimental investigation.

#### **3.5.4 A Trade-Off between Post-Stop F0 and Tone?**

The fact that the post-stop F0 weight is diminished in the Mandarin context across both production and perception raises the question of whether the lost weight in post-stop F0 is transferred to other dimensions, with the most obvious candidate being tonal category. In what follows, I discuss the case with production first before moving on to perception.

The question about the existence of a trade-off between post-stop F0 and tone is tied to the debate of whether tone attenuates the degree of post-stop F0 difference. As mentioned in Section 3.2.2, whereas there are some studies that point to a positive direction (e.g., Gandour, 1974; Hombert, 1978), large magnitudes of post-stop F0 difference have also been observed in tonal languages (e.g., Francis et al., 2006; Phuong, 1981; Shimizu, 1994; Xu and Xu, 2003). In the current study, the Mandarin-English bilinguals' respective language productions do conform to the former pattern at the population level. However, not every speaker matches the population-level trend, with some speakers producing the post-stop F0 effect

to a similar degree in both languages. The results presented here thus agree with Kirby's (2018) observation that attenuation of post-stop F0 effect in tone languages depends on speaker-specific implementation of laryngeal maneuvers to distinguish voicing and tone.

With respect to perception, as described in Section 3.2.6, it is possible that, in interpreting the audio stimuli as Mandarin words, Mandarin-English bilinguals attribute the variation in post-stop F0 partially to the lexical tones in the language, and that in treating the stimuli as English words, they ascribe the variation to stop voicing. If this is indeed the case, then it is expected that the post-stop F0 weight would be correlated with the tone weight across the two languages. That is, the decrease in post-stop F0 weight from English to Mandarin would be accompanied by an increase in tone weight. Looking at Table 3.11, which shows the results at the population level, it seems the loss in post-stop F0 is indeed accompanied by an increase in tone weight, though the model is not as confident in the increase in tone weight as in the decrease in post-stop F0 weight. At the individual level, the panels for post-stop F0 and tone in Figure 3.17 also appear to suggest that for many participants, a drop in post-stop F0 weight is compensated by a rise in tone weight, and that those who have a bigger drop tend to have a sharper rise as well (notice the apparent negative correlation in Figure 3.17I when the changes along these two dimensions are plotted against each other), at least as far as the posterior mean is concerned. However, the correlation coefficient estimated from posterior samples does not back up this hypothesis (as shown by the lines going into different directions in Figure 3.17I). Therefore, it is still inconclusive as to whether there is a trade-off relation between post-stop F0 and tone in perception.

### **3.5.5 Production-Perception Interface**

As shown in Section 3.4.8, even though the use of post-stop F0 is mirrored across production and perception at the macro level, the link between the two modalities at the individual level turns out to be tenuous, at least for Mandarin. Similar null results have been reported for VOT and F0 in English (Shultz et al., 2012), VOT, F0, closure duration, and F1 onset for English and Spanish (Schertz, 2014), or VOT, F0, and closure duration in L1 Korean and L2 English (Schertz et al., 2015),

among other studies that used fairly standard paradigms similar to the one employed in this study. It seems that a lack of relationship is the norm in studies that sought to establish individual-level correlation in cue use. Given that population-level correspondences between the modalities alone cannot be taken as evidence for a causal link—if there is a (direct or indirect) causal link between the modalities, it should surface on an individual level (Schertz et al., 2020)—the patterns in the data here therefore demonstrate partial independence between perception and production, at least when characterizing post-stop F0 weights in Cohen’s  $d$  and logistic coefficients. This lack of a direct production-perception link also casts doubt on whether perceptual cue weights can be derived from distributional information in production alone, as is posited in some models, such as Motor Theory (Lieberman and Mattingly, 1985) or Direct Realism (Fowler, 1986). For instance, during a sound change in progress, some speakers tend to rely more on the “innovative” cue and others on the “traditional” cue. In order to understand speakers at different stages of the change, an listener’s perception needs to be more flexible and therefore cannot strictly echo their own production (Beddor et al., 2018).

### **3.6 Conclusion**

The current work examined whether and how L1 Mandarin-L2 English bilinguals use post-stop F0 as a cue for stop voicing across production and perception in Mandarin as well as English contexts. The production results show that F0 is actively used to encode both tonal and voicing distinctions in their Mandarin tokens, and that voicing distinctions are likewise embedded with post-stop F0 in English tokens. In perception, the bilinguals are also able to extract voicing information from post-stop F0 (in the same direction as observed in production) in both languages, even when post-stop F0 is integrated in the overall pitch contour, which they need to monitor in order to identify the lexical tone. Crucially, the reliability of post-stop F0 in signaling the voicing contrast and the extent to which the bilinguals lean on post-stop F0 for voicing perceptually are language-specific, such that production and perceptual weights for post-stop F0 are greater in the English context. However, lack of correlation between individual’s use of post-stop F0 across production and perception suggests that a strict view that the variability in percep-

tual cue weights is reflective of individual differences in production patterns might be incorrect.

## **Chapter 4**

# **Differences in Post-Stop F<sub>0</sub> in the Production and Perception of the English Stop Voicing Contrasts by L1 English and L1 Mandarin Speakers**

### **4.1 Introduction**

In contrast to the previous chapter, which focuses on L1 (first language) Mandarin-L2 (second language) English bilinguals' production and perception of the stop voicing contrasts in Mandarin and English, this chapter compares L1 Mandarin-L2 English bilinguals' (or L1 Mandarin for short) production and perception of the stop voicing contrasts in English with those from L1 English speakers. The results from such a comparison can shed light on the interplay between cue primacy and L1's influences on L2.

This study treats L1 English speakers' performances as the reference against which L1 Mandarin speakers' English performances are compared. Given that the L1 Mandarin participants in this study speak English as an L2, are of simi-

lar ages, and all grew up in China, any divergence between their and L1 English speakers' performances is most likely to be driven by the influence from Mandarin. Against this background, this section first reviews two major theories of L2 speech sound acquisition—the Speech Learning Model (SLM, Flege, 1995, 2007) and the Perceptual Assimilation Model's extension to L2 acquisition (PAM-L2, Best and Tyler, 2007)—with a focus on their predictions for behavior in the production and perceptual tasks described in Section 3.3 and Section 3.4. In addition, as Mandarin differs from English in that fundamental frequency (F0) is used to contrast lexical items in the former, the perceptual consequences of this functional dichotomy of F0 in terms of its processing interference with other contrastive units, such as consonants, are discussed. Finally, the research questions and predicted results based on aforementioned theories are laid out.

#### **4.1.1 Speech Learning Model (SLM)**

Beginning with the postulate that the capacities to learn the language-specific properties of L1 remain unchanged across the life span and remain accessible to L2 learners of all ages—SLM explains the mutual influence between L1 and L2 by positing that the phonetic categories (i.e., language-specific aspects of speech sounds specified in long-term memory representations) making up the L1 and L2 subsystems of a bilingual operate in a “common phonological space”. In particular, two mechanisms—phonetic category assimilation and phonetic category dissimilation—are proposed, through which the phonetic categories comprising the L1 and L2 phonetic subsystems interact, and which of the two mechanisms applies depends on whether category formation has taken place for L2 sounds or not.

Whether a new phonetic category will be established for an L2 sound in turn hinges on the perceived phonetic (dis)similarity of the L2 sound from the closest L1 sound. If an L2 sound is perceived to be phonetically similar to an L1 sound, and neighboring L1 categories are fully developed, equivalence classification with an L1 sound (i.e., an L2 sound is processed as an instance of an L1 category) is most likely, and a new category for the L2 sound will be blocked. However, as repeatedly emphasized in the framework, equivalence classification does not

prevent L2 learners from auditorily detecting cross-language phonetic differences. In fact, the SLM predicts that production of an L2 sound will be modified slowly over time if the L2 sound differs audibly from the L1 sound with which it has been equated. This modification might even result in a “merged” category that takes on both L1 and L2 phonetic properties, such that this merged category might possess phonetic properties that are different from monolinguals of each language. Such a merging scenario is illustrated in Figure 4.1A, where the English /b/ is equivalence-classified as the Mandarin /p/ and merged into that category. Under the formation of a merged category is the implicit claim of the SLM that perception *leads* production, so production and perception become more aligned with each other over the course of learning.

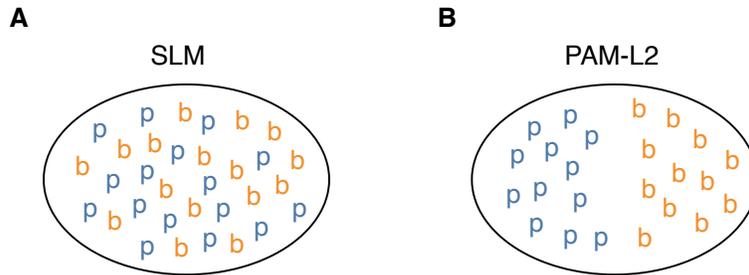
On the other hand, if bilinguals discern at least some of the phonetic differences between the L1 and L2 sounds, a new phonetic category might be established for an L2 sound. The newly-established phonetic category will make the phonetic space of the L2 learner more crowded, which then might push the L1 and L2 phonetic categories away from one another in order to maintain contrast. If this happens, the newly-formed category might dissimilate from a neighboring L1 category, or vice versa, causing a bilingual’s productions of the L2 and/or the L1 sounds to be different from those of monolinguals. Therefore, in the case of category merger or dissimilation, the SLM predicts that the end result will be difference from monolinguals.

#### **4.1.2 Perceptual Assimilation Model (PAM)**

PAM was originally developed to explain nonnative speech perception by naive listeners (Best, 1995) but has been extended to PAM-L2, which addresses the issues of L2 learners’ perceptual difficulties and biases across variations or contrasts in target languages (Best and Tyler, 2007). PAM-L2 is founded on the ecological direct-realist premise that the focus of speech perception is on information about the distal articulatory events that produced the speech signal (e.g., Best, 1995). Although the model itself does *not* explicitly make claims about *production* of novel contrasts, the model’s basis in the Direct Realism (Fowler, 1986) implies shared representations between speech production and perception. One can there-

fore infer that learning in one modality is strongly correlated to learning in the other modality. This view that perceivers extract invariants about articulatory gestures from the speech signal therefore contrasts with the SLM, which underscores the forming of categories from acoustic-phonetic cues. In fact, PAM-L2 posits that language-relevant speech properties are differentiated not only at the phonetic level but also at the higher-order phonological level, as well as at the lower-order gesture level. L1-L2 differences at a gestural, phonetic, or phonological level can therefore all potentially influence the L2 learner's discrimination abilities. In the PAM-L2 framework, it is at the phonological level that listeners may identify L1 and L2 sounds as functionally equivalent. For instance, English L2 learner of French tend to equate the category /r/ across the two languages because the two phonemes reflect a similar patterning of rhotics across the two languages in syllable structure, phonotactic regularities, and allophonic alternations (Ladefoged and Maddieson, 1996; Lindau, 1985), despite the fact that the two phonemes bear relatively little phonetic similarity to each other (i.e., [ʁ] for French /r/ and [ɹ] for English /r/). That is, different phonetic realizations may be learned for an equated phonological category across languages. The possibility for L1 and L2 to interact at both phonetic and phonological levels leads to a different prediction from that of the SLM: even when an L2 phone is treated as belonging to the same phonological category of an L1 phone, if the phonetic difference between the two phones is audible, the listener should still be able to maintain the L1 and L2 phones as separate phonetic realizations of the one phonological category. This case is schematized in Figure 4.1B, where the English /b/ is treated as belonging to the same phonological category of the Mandarin /p/, but the two phones are maintained as separate phonetic realizations.

Both the SLM and PAM-L2 relate the difficulties an L2 learner might face in acquiring the phonological/phonetic categories of L2 to the cross-linguistic differences between the L1 and L2 sound systems. However, another way L2 learners might process the speech signals in the target language differently than L1 listeners lies in the language-specific interactions between various linguistic units in their L1 (e.g., Choi et al., 2019; Kim and Tremblay, 2021). As the present study concerns L1 listeners of Mandarin, which is a tonal language, the next section reviews how the processing of lexical tone influences and is influenced by the processing



**Figure 4.1:** Illustration of equivalence classification between an L1 sound (i.e., /p/ in Mandarin) and an L2 sound (i.e., /b/ in English) that are phonetically similar but still auditorily different. **A.** The SLM predicts an eventual merged category that takes on both L1 and L2 phonetic properties. **B.** The PAM-L2 allows for a single phonological category with different language-specific phonetic realizations.

of consonant, and how such a processing dependency is carried over to their L2 English.

### 4.1.3 Perceptual Interference between Consonants and Tone

While lexical tones and consonants are usually treated as independent units in phonology, the actual speech signal produced by speakers simultaneously carries information about consonants (and vowels) and lexical tones, and listeners also have to extract the two types of information from the acoustic waveform during speech processing. One perceptual consequence of this multidimensionality of speech sounds is that there are perceptual dependencies (which is also referred to as *mutual integrality* or *dimensional integrality*) between segmental and suprasegmental dimensions, which has typically been studied using the Garner speeded classification paradigm (Garner, 1974, 1976; Garner and Felfoldy, 1970). The paradigm involves two prototypical tasks—baseline and orthogonal. In the baseline task, participants are asked to make speeded classification judgments according to the target dimensions (e.g., voice onset time [VOT]) while the non-target dimension (e.g., F0 contour on the vowel) is held constant. In the orthogonal task, the task is identical except that the values along the non-target dimension also vary on a trial-to-trial basis. As such, change in performance between the orthogonal

and baseline tasks must originate from the variability in the non-target dimension, and such a difference in performance between tasks is referred to as Garner interference. For instance, Tong et al. (2008) use the Garner speeded classification paradigm to examine processing interactions between segmental (i.e., consonant and vowel) and suprasegmental (tone) dimensions of Mandarin in L1 Mandarin listeners. In their experiment, listeners are tasked to attend to variation along one dimension (consonant: /b/ vs. /d/; vowel: /a/ vs. /u/; tone: Tone 2 [high-rising] vs. Tone 4 [high-falling]) while ignoring the variation along another. Their results show that interference effects, as measured by  $d'$  (which is in turn based on accuracy) and response time difference across the orthogonal and baseline conditions, are asymmetric between segmental and suprasegmental dimensions, with the segmental dimension interfering more with tone classification than the other way around.

Processing dependencies between consonantal and tonal information are not only observed for listeners of tonal languages. Earlier cross-linguistic studies of Mandarin with the Garner paradigm suggest that *both* Mandarin and English listeners show mutual integrality between consonants and tone (Lee and Nusbaum, 1993; Repp and Lin, 1990). For instance, in Lee and Nusbaum's (1993) perception experiment with the Garner paradigm, L1 English and L1 Mandarin listeners heard CV syllables varying on a consonantal dimension (/b/ vs. /d/) and either a Mandarin lexical tonal dimension (Tone 3 [low-dipping] vs. Tone 4 [high-falling]) or a non-Mandarin constant-pitch suprasegmental dimension (low [104 Hz] vs. high [140 Hz]). Both groups had to respond along a given dimension (*b* or *d*, *3rd* or *4th*, *high* or *low*) by pressing one of the designated keys on a keyboard. They found that L1 Mandarin listeners processed both the consonantal and suprasegmental dimensions with mutual integrality for both the Mandarin and the constant-pitch stimuli, whereas L1 English listeners showed mutual integrality only with the Mandarin stimuli. In interpreting this finding, however, it should be noted that the relative discriminability between dimensions can affect the degree of interference of one dimension on the other (e.g., classification along the less discriminable dimension, either due to acoustic distance or degree of familiarity, will be affected by variation along the more discriminable dimension to a greater extent than vice versa). As previous studies involving lexical tone did not control for discriminability of the

tonal dimension for English and Mandarin listeners, cross-linguistic comparisons of the dimensional integrality between consonant and tone might not be conclusive as to whether the strength of dimensional interaction is indeed the same for both groups of listeners. However, given the fact that Mandarin listeners need to actively track tonal information in order to identify lexical items, it is expected that, to the extent that segmental and tonal dimensions are interacting with each other perceptually, the interference from the tonal dimension when L1 Mandarin listeners are processing segmental information would be more severe as compared to L1 English listeners.

Another piece of evidence for tonal information being constantly integrated in L1 Mandarin listeners' speech processing comes from a study by Shook and Marian (2016). In their eye-tracking experiment, L1 Mandarin-L2 English bilinguals hear an English word and have to select which of two visually presented Chinese characters corresponds to the correct Mandarin translation. The crucial manipulation involves the pitch contour of the spoken (English) word: it either matches or mismatches the lexical tone of the Mandarin translation. Their results show that bilinguals are faster to identify the correct translation and make earlier eye movements to the translation when the pitch contour of the spoken word matches that of its Mandarin translation. The authors interpret the results as indicating a high degree of interactivity between the participants' L1 Mandarin and L2 English. In the context of the present study, since the audio stimuli also vary in the pitch contour that either matches Mandarin Tone 1 (high-level) or Tone 4 (high-falling), in addition to VOT and post-stop F0, this additional variation in the pitch contour, which is irrelevant to the identification of English lexical items in the reported perception experiment below, might cause L1 Mandarin listeners' performance to diverge from that of L1 English listeners, as expounded in the following section.

#### **4.1.4 Goals of the Current Study**

Given that Mandarin uses F0 extensively for its lexical tone contrast, it provides an opportunity to examine L1 influence on L2 cue weighting from a new perspective: how do L1 tonal language speakers make use of the post-stop F0 cue when producing and perceiving stop voicing contrasts in their non-tonal L2, in comparison with

native speakers of the L2? In this case study, the L1 tonal language is exemplified by Mandarin while the L2 non-tonal language is represented by English. The production experiment compares the usage patterns of post-stop F0 by L1 Mandarin speakers in contrasting the stop voicing in their L2 English with those of L1 English speakers. The perception experiment turns to the perception of English stops in terms of the same acoustic dimension: the same group of L1 Mandarin and L1 English listeners participated in a four forced-choice identification task for stimuli covarying in VOT, post-stop F0, and pitch contour.

The expected outcomes derived from the frameworks and observations described in the previous sections are summarized in Table 4.1 and dissected below.

**Table 4.1:** Predicted production and perception results under different frameworks.

<b>Framework</b>	<b>Predicted production results</b>	<b>Predicted perception results</b>
SLM and PAM-L2	L1 Mandarin speakers are predicted to use the post-stop F0 cue to a lesser degree than L1 English speakers, because of the influence from L1 Mandarin. In addition, the two frameworks predict that production should mirror perception, such that if L1 Mandarin and L1 English speakers use post-stop F0 to a different degree in production, the two groups should use post-stop F0 to a different extent in perception.	
Perceptual interference	—	Post-stop F0 for L1 Mandarin listeners has a lower weight and/or more variability

### **Predictions from the SLM and PAM-L2**

The expected outcomes of the production and perception experiments can be explored through the SLM and PAM-L2. Even though the SLM and PAM-L2 differ from each other mechanically in explaining the assimilation/dissimilation patterns in L2 sound acquisition, the two frameworks make very similar predictions, as far as the current study is concerned, so the predictions from the two frameworks will be collapsed in the following discussion. Applying the SLM and PAM-L2 to L1 Mandarin speakers' production and perception of post-stop F0 in the English stop

voicing contrast, two diverging outcomes are anticipated, depending on whether new phonetic categories are established for L2 English stops.

Given that both English and Mandarin have two stop categories with respect to phonological voicing (voiceless aspirated vs. voiceless unaspirated for Mandarin, and voiceless vs. voiced for English), and that the voicing contrast of stops in both languages make use of VOT (i.e., the absence/presence of aspiration) as the primary cue, English phonemically voiced (/b, d, g/) and voiceless (/p, t, k/) stops in the word-initial position will almost certainly be assimilated to Mandarin unaspirated (/p, t, k/) and aspirated stops (/p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>/) respectively (e.g., Chen et al., 2008; Ding et al., 2018). In this scenario, no new L2 sound categories need to be established, meaning that L2 English voiced (voiceless) stops will be processed as instances of L1 Mandarin unaspirated (aspirated) stops. In addition, as reviewed in Section 3.2.2 in the previous chapter, the post-stop F0 difference between voiceless and voiced stops in English is larger than that between aspirated and unaspirated stops in Mandarin. If we further assume that L1 Mandarin listeners are able to detect this cross-linguistic post-stop F0 difference, in the case where new phonetic categories do not take shape for L2 English stops, the SLM predicts a merged category that has a post-stop F0 value in between the typical values in Mandarin and English (Figure 4.1A). However, the results from Chapter 3 have shown that L1 Mandarin speakers change post-stop F0 difference in response to the current language context, so the prediction of a merged category is not borne out. On the other hand, PAM-L2 predicts that, even when the Mandarin and English stops are treated as a single phonological category, bilinguals might still produce and perceive the subtle differences in the post-stop F0 across the two languages (Figure 4.1B). That is, we would expect L1 Mandarin speakers to differentiate post-stop F0 in their production according to the language context and to also rely on post-stop F0 to different degrees in perception across languages. Indeed, these predicted patterns agree with the experiment results reported in the previous chapter, at least as far as post-stop F0 is concerned. With regard to the comparison of the use of post-stop F0 in English by L1 Mandarin and L1 English speakers, since English stops are highly likely to be assimilated into their Mandarin counterparts, it is expected that L1 Mandarin speakers' performance would be influenced by their native language to some degree. Specifically, even though the results from the experiments in Chap-

ter 3 demonstrate that L1 Mandarin speakers were able to adapt their post-stop F0 weights to the language context, the fact that they are likely to process similar Mandarin and English stops as a single category means that their English performance might still be different from that of L1 English speakers. In particular, the extent to which L1 Mandarin speakers rely on post-stop F0 is predicted to be generally lower than that of L1 English speakers. Additionally, as is implicitly assumed in the SLM and explicitly stated in the PAM-L2, production and perception are expected to be tightly connected in both frameworks, such that if L1 Mandarin speakers perform differently from L1 English speakers in production, then the two groups should behave differently in perception as well. The relationship between participants' production and perception of post-stop F0 in the context of stop contrasts is further explored through correlation analyses to determine whether the reliability of post-stop F0 in a speaker's production can predict the same participant's reliance on post-stop F0 in perception.

#### **Predictions from Perceptual Interference between Consonants and Tone**

More specific predictions on the perceptual front can be made when taking into account the processing dependency between consonants and tone. If it is indeed the case that L1 Mandarin listeners process segmental and tonal information in a more integrated way, and that they are sensitive to irrelevant tonal variation, then we would expect L1 English listeners and L1 Mandarin listeners to differ with respect to the use of the post-stop F0 cue. In particular, given that the manipulation of post-stop F0 is piggybacked on lexical tone, and that L1 Mandarin listeners are likely to pay attention to tonal variation even when this information is not needed for the identification task, changes in post-stop F0 might be perceived as part of the tone, rendering post-stop F0 an unreliable cue for stop voicing for L1 Mandarin listeners. Quantitatively, post-stop F0 would therefore be assigned a lower weight and/or have more variability, both within-subject and between-subject, for L1 Mandarin listeners.

## 4.2 Production Experiment

This experiment compared L1 English and L1 Mandarin speakers' productions of post-stop F0 in English. The L1 Mandarin speakers are the same participants taking part in the experiments described in Chapter 3, and their production data from the English production experiment in Chapter 3 is recycled here to form new comparisons with the production of L1 English speakers.

### 4.2.1 Participants

Two groups of speakers took part in the experiment: 25 L1 English speakers (17 female, 8 male;  $\text{mean}_{\text{age}} = 25.3$  yrs,  $\text{median}_{\text{age}} = 22$  yrs,  $\text{range}_{\text{age}} = [19, 47]$  yrs,  $\text{SD}_{\text{age}} = 8.6$  yrs) and 25 L1 Mandarin-L2 English speakers. The 25 L1 Mandarin-L2 English speakers are the same as those participating in the experiments described in Chapter 3. All L1 English participants were recruited from the linguistic participant pools at either the University of British Columbia or the University of Toronto, and received partial course credit as compensation. The demographic information of the included L1 English participants is given in Table E.1 in the appendix. For L1 English participants' data to be considered for analyses, they must fulfill the following requirements, which are analogous to those of the Mandarin ones in Chapter 3:

1. They completed all required experiment components;
2. They self-report as a native speaker of English;
3. They have at least one primary caretaker whose native language is English;
4. They do not speak any tonal languages fluently (i.e., they rate their speaking ability below 3 on a 0-6 scale or below "fair") but could be speakers of other non-tonal languages;
5. They lived in an English-speaking country for at least 10 years between birth and age 15.

In addition, as evaluated by the author, participants whose entire recordings contained excessive background noise due to their doing the experiment in a noisy

place, or were extremely quiet that made it challenging to identify acoustic landmarks for annotation, were omitted from the analyses altogether.

#### **4.2.2 Stimuli, Procedure, Annotations, and Measurements**

The stimuli, procedure, recording annotations, and acoustic measurements were identical to those of the English version of the production experiment, which was described in Section 3.3 in Chapter 3.

#### **4.2.3 Omitted Data**

Tokens that had the following problems were excluded from analyses entirely: mispronunciations, as judged by the author (1 from L1 English, 26 from L1 Mandarin), skipped tokens (3 from L1 Mandarin), technical issues (3 from L1 English, 4 from L1 Mandarin), and creaky vowel onset, as determined from the spectrogram and impressionistically by the author (36 from L1 English, 33 from L1 Mandarin).

#### **4.2.4 Statistical Analyses**

Two sets of models were involved, similar to the analyses presented in Section 3.3.7. The first set examined whether vowel-onset F0 had different values following different onset types in L1 English and L1 Mandarin speakers' productions, and the second set focused on the computation of production weight for post-stop F0 for both speaker groups.

For the first set of models, vowel-onset F0 was first  $z$ -transformed within each speaker. Similar to Chapter 3, the choice of  $z$ -transformation was informed by the discussion presented in Chapter 2 that  $z$ -transformation preserves the relative ordering of tokens in terms of raw vowel-onset F0, and that it facilitates the specification of priors. The model consisted of a linear submodel predicting vowel-onset F0 in the English words produced by L1 English speakers, a linear submodel predicting vowel-onset F0 in the tokens produced by L1 Mandarin speakers, and terms connecting the two submodels. The two submodels included the same population-level terms, which came from the combinations of the following variables: the **voicing** of the onset consonant (ASPIRATED, UNASPIRATED,

SONORANT;<sup>1</sup> forward-difference-coded with ASPIRATED vs. UNASPIRATED and UNASPIRATED vs. SONORANT), the **height** of the main vowel (HIGH [i], LOW [aɪ]; sum-coded with LOW being coded with  $-1$  and HIGH with  $1$ ), and consonantal **places of articulation (PoA)**; BILABIAL, ALVEOLAR, VELAR; sum-coded with BILABIAL being coded with  $-1$ ). To account for how each predictor affected the realization of the voicing contrast, two-way interaction terms between **voicing** and the other predictors were also considered in the model comparison process, provided that the main effect of the predictor was included.

The two submodels were tied together via a two-level hierarchical structure: the first level estimated the language-group-specific variation while the second level estimated the speaker-specific variation within each language group. Accordingly, each coefficient in the model had three parts: a language-universal estimate, a language-group-specific adjustment, and a speaker-specific adjustment. Language-universal estimates all had a regularizing prior of  $\text{Normal}(\mu = 0, \sigma = 5)$ . Language-group-specific adjustments were assumed to have a diagonal covariation structure, so, for example, the intercept adjustment for L1 English speakers was independent of the height adjustment for the same group of speakers. Correlations among speaker-specific adjustments used the Lewandowski-Kurowicka-Joe (LKJ) prior with  $\xi = 1$  (Bürkner, 2017). Finally,  $\text{Exponential}(r = 1)$  was used as the prior for the error term, as well as for other variational parameters needed in the hierarchical structure. The formal specification for the model whose results were reported is given in Section E.2.1 in the appendix.

Again, evidence embedded in the model was evaluated by the posterior distributions of coefficients/parameters and by model comparison. The detailed procedure can be found in Section 2.2.3. Candidate models were constructed incrementally, using both prior knowledge and model comparison results. All the candidate models are given in Table 4.2. The base model (i.e., M1) included only **height**, given that vowel height is known to affect F0 (Whalen and Levitt, 1995). Other factors and their interactions, such as **voicing** and **PoA**, were then gradually added to the base model.

The second set of analyses focused on the quantification of post-stop F0 weight

---

<sup>1</sup>Following the convention from the previous chapter, I continue to use the labels *aspirated* and *unaspirated* stops to refer to the (phonologically) voiceless and voiced stops in English, respectively.

**Table 4.2:** Candidate vowel-onset F0 models considered in the model comparison, with their ELPD-LOO means and standard deviations. An intercept was included in each model but is omitted here.

<b>Model</b>	<b>ELPD-LOO</b>	<b>ELPD-LOO standard error</b>	<b>Predictors</b>
M1	−3445.8	43.5	height
M2 (final)	−3053.7	44.5	height + voicing
M3	−3041.4	44.8	height + voicing + PoA
M4	−3045.9	45.0	height + voicing + voicing × height

for L1 English and L1 Mandarin speakers. As in Chapter 3, the production weight of a cue aims to measure how reliable the cue is in separating different members of a phonological contrast, and was operationalized based on the amount of overlap between the categories along the acoustic dimension in question. Cohen’s  $d$  was again used as the proxy for production weight, so a higher weight means post-stop F0 is more reliable in separating aspirated tokens from unaspirated ones. The structure of the model used to derive Cohen’s  $d$  at the population and individual levels was very similar to the one described in Section 3.3.7. Specifically, for each language group, a Bayesian mixed model was fitted to estimate the means and standard deviations of the post-stop F0 distributions of the aspirated and unaspirated categories, both at the population and individual levels. These estimated means and standard deviations were then used to infer production weights for the whole speaker group and for individual speakers. Since estimated means and standard deviations were distributions, each derived weight therefore also formed a distribution.

#### 4.2.5 Results

Mean production values and standard deviations for L1 English and L1 Mandarin speakers’ VOT and post-stop F0 are given in Table 4.3. Distributions of standardized vowel-onset F0 values, broken down by the onset type, vowel height, place of articulation, and speaker group, are plotted in Figure 4.2. Table 4.2 shows the ELPD-LOO mean and standard error for each candidate model, and Table 4.4 gives

model comparison results in terms of ELPD-LOO differences and associated standard errors.

**Table 4.3:** Means and standard deviations for VOT and vowel-onset F0 in hertz (split by gender) from L1 English and L1 Mandarin speakers’ productions of English word-initial stops and sonorants.

	<b>L1 English speaker</b>			
	<b>Aspirated</b>	<b>Unaspirated (short-lag)</b>	<b>Unaspirated (lead)</b>	<b>Sonorant</b>
VOT (ms)	98 (29) <i>n</i> = 438	23 (14) <i>n</i> = 331	−101 (40) <i>n</i> = 164	— —
F0, male (Hz)	130 (21) <i>n</i> = 139	122 (19) <i>n</i> = 79	109 (13) <i>n</i> = 80	112 (17) <i>n</i> = 143
F0, female (Hz)	240 (26) <i>n</i> = 299	225 (30) <i>n</i> = 252	224 (22) <i>n</i> = 84	211 (21) <i>n</i> = 303

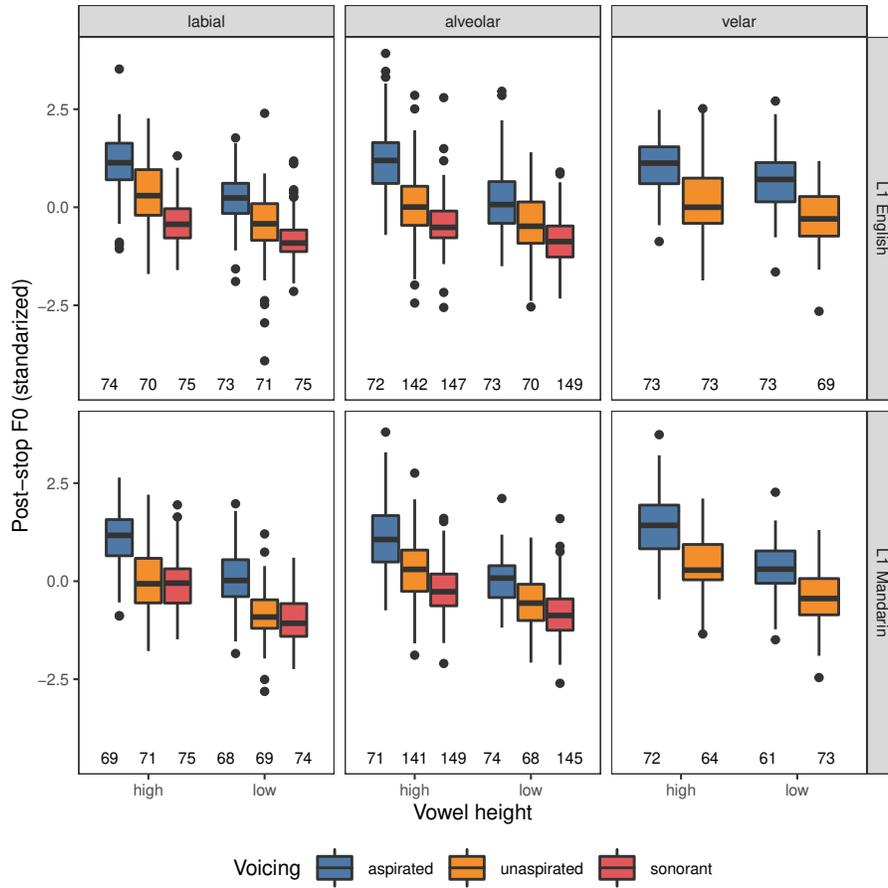
  

	<b>L1 Mandarin speaker</b>			
	<b>Aspirated</b>	<b>Unaspirated (short-lag)</b>	<b>Unaspirated (lead)</b>	<b>Sonorant</b>
VOT (ms)	111 (32) <i>n</i> = 415	21 (12) <i>n</i> = 446	−143 (62) <i>n</i> = 40	— —
F0, male (Hz)	146 (20) <i>n</i> = 178	138 (18) <i>n</i> = 178	120 (18) <i>n</i> = 40	131 (19) <i>n</i> = 193
F0, female (Hz)	265 (28) <i>n</i> = 237	248 (30) <i>n</i> = 268	— —	237 (26) <i>n</i> = 250

**Table 4.4:** Model comparison results for key model pairs, expressed in differences in ELPD-LOO and associated standard errors. Pairs that are judged to differ in predictive power are marked by asterisks.

<b>Model</b>	<b>M2</b>	<b>M3</b>	<b>M4</b>
<b>M1</b>	−392.1 (34.0)*		
<b>M2</b>		−12.3 (7.0)	−7.8 (5.8)

The results of model comparison corroborated the importance of phonological voicing in predicting vowel-onset F0 (i.e., M1 vs. M2); however, adding place of



**Figure 4.2:** Boxplot of standardized post-stop F0 values, normed by speaker, as a function of place of articulation, language group, vowel height, and phonological voicing.

articulation and the interaction between voicing and vowel height did not seem to improve the predictive power much (i.e., M2 vs. M3 and M2 vs. M4). Therefore, the most parsimonious model M2 was selected as the final model. In presenting the results from the model, those at the population level are described prior to the results at the individual level.

### **L1 English and L1 Mandarin Speakers' English Production: Population Results**

The marginal posterior distributions of population-level parameters from M2 are summarized in Table 4.5. As anticipated, the high vowel /i/ tended to induce a higher onset F0 for both groups of speakers (L1 English:  $\bar{\beta} = .30$ , 89% CrI = [.25, .35],  $p(\beta > 0) = 1.00$ ; L1 Mandarin:  $\bar{\beta} = .43$ , 89% CrI = [.37, .48],  $p(\beta > 0) = 1.00$ ). When comparing aspirated stops with unaspirated stops, the former on average was associated with a higher post-stop F0 for both speaker groups (L1 English:  $\bar{\beta} = .87$ , 89% CrI = [.73, 1.03],  $p(\beta > 0) = 1.00$ ; L1 Mandarin:  $\bar{\beta} = .83$ , 89% CrI = [.70, .96],  $p(\beta > 0) = 1.00$ ). Unaspirated stops in turn led to a higher onset F0 than sonorants, again, for both speaker groups (L1 English:  $\bar{\beta} = .47$ , 89% CrI = [.32, .63],  $p(\beta > 0) = 1.00$ ; L1 Mandarin:  $\bar{\beta} = .36$ , 89% CrI = [.27, .44],  $p(\beta > 0) = 1.00$ ). Crucially, if the magnitude of post-stop F0 difference due to voicing is contrasted across speaker groups, the two groups were comparable, with respect to both the difference between aspirated and unaspirated categories, and that between unaspirated and sonorant categories (ASPIRATED vs. UNASPIRATED:  $\bar{\beta} = .04$ , 89% CrI = [-.14, .21],  $p(\beta > 0) = .65$ ; UNASPIRATED vs. SONORANT:  $\bar{\beta} = .11$ , 89% CrI = [-.06, .27],  $p(\beta > 0) = .84$ ).

### **L1 English and L1 Mandarin Speakers' English Production: Individual Results**

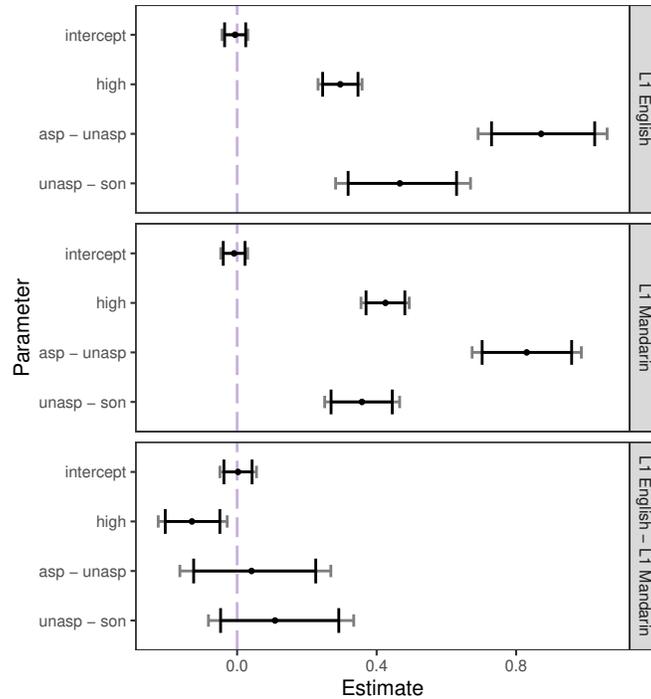
The distributions for parameters from M2 at the individual level are plotted in Figure 4.4; the numerical values associated with the summary plot can be found in Table E.2 and Table E.3 in the appendix.

As the focus of this study is on the relationship between post-stop F0 and onset voicing, the following discussion is centered on the effect voicing has on post-stop F0 and ignores the effect associated with vowel height. For both L1 English and L1 Mandarin speakers, almost all speakers (i.e., 24/25 in L1 English and 24/25 in L1 Mandarin) had a robust positive difference between post-stop F0 following an aspirated stop and that following an unaspirated stop. However, the post-stop F0 difference between an unaspirated stop and a sonorant paints a slightly different picture: even though the results at the population level suggested both groups are comparable in terms of the magnitude of onset F0 difference between

**Table 4.5:** Marginal posterior summaries for key parameters from M2. The contrast coding scheme for each variable is explained in Section 4.2.4. The parameters whose effects are judged to be strong are marked with \*\*, and those whose effects are judged to be weak are marked with \*.

Parameter	Mean	SD	89% CrI	$p(\text{dir.})$
intercept: ES	-.01	.02	[-.04, .02]	$p(\beta < 0) = .62$
high – (high + low)/2: ES**	.30	.03	[.25, .35]	$p(\beta > 0) = 1.00$
asp – unaspl: ES**	.87	.09	[.73, 1.03]	$p(\beta > 0) = 1.00$
unaspl – son: ES**	.47	.10	[.32, .63]	$p(\beta > 0) = 1.00$
intercept: MS	-.01	.02	[-.04, .02]	$p(\beta < 0) = .68$
high – (high + low)/2: MS**	.43	.04	[.37, .48]	$p(\beta > 0) = 1.00$
asp – unaspl: MS**	.83	.08	[.70, .96]	$p(\beta > 0) = 1.00$
unaspl – son: MS**	.36	.05	[.27, .44]	$p(\beta > 0) = 1.00$
intercept: ES – MS	.00	.02	[-.04, .04]	$p(\beta > 0) = .55$
high – (high + low)/2: ES – MS**	-.13	.05	[-.21, -.05]	$p(\beta < 0) = .99$
asp – unaspl: ES – MS	.04	.11	[-.14, .21]	$p(\beta > 0) = .65$
unaspl – son: ES – MS	.11	.11	[-.06, .27]	$p(\beta > 0) = .84$

unaspirated and sonorant series, the individual results revealed that there was more variation in L1 English speakers' production. In particular, while the model indicated that all L1 Mandarin speakers reliably produced a raised onset F0 after an unaspirated stop, there were a number of L1 English speakers for whom the model had less evidence for a consistent post-stop F0 difference between the two consonant series. Interestingly, upon careful inspection of Figure 4.4, there seemed to be a negative correlation between the amount of post-stop F0 difference across the aspirated and unaspirated categories and that across the unaspirated and sonorant categories for L1 English speakers (Figure 4.4A), but not for L1 Mandarin speakers (Figure 4.4B). This observation is backed up by model output: 89% CrI of  $\rho_{\text{asp} - \text{unaspl}, \text{unaspl} - \text{son}} = [-.88, -.46]$  for L1 English speakers and 89% CrI of  $\rho_{\text{asp} - \text{unaspl}, \text{unaspl} - \text{son}} = [-.73, .48]$  for L1 Mandarin speakers.

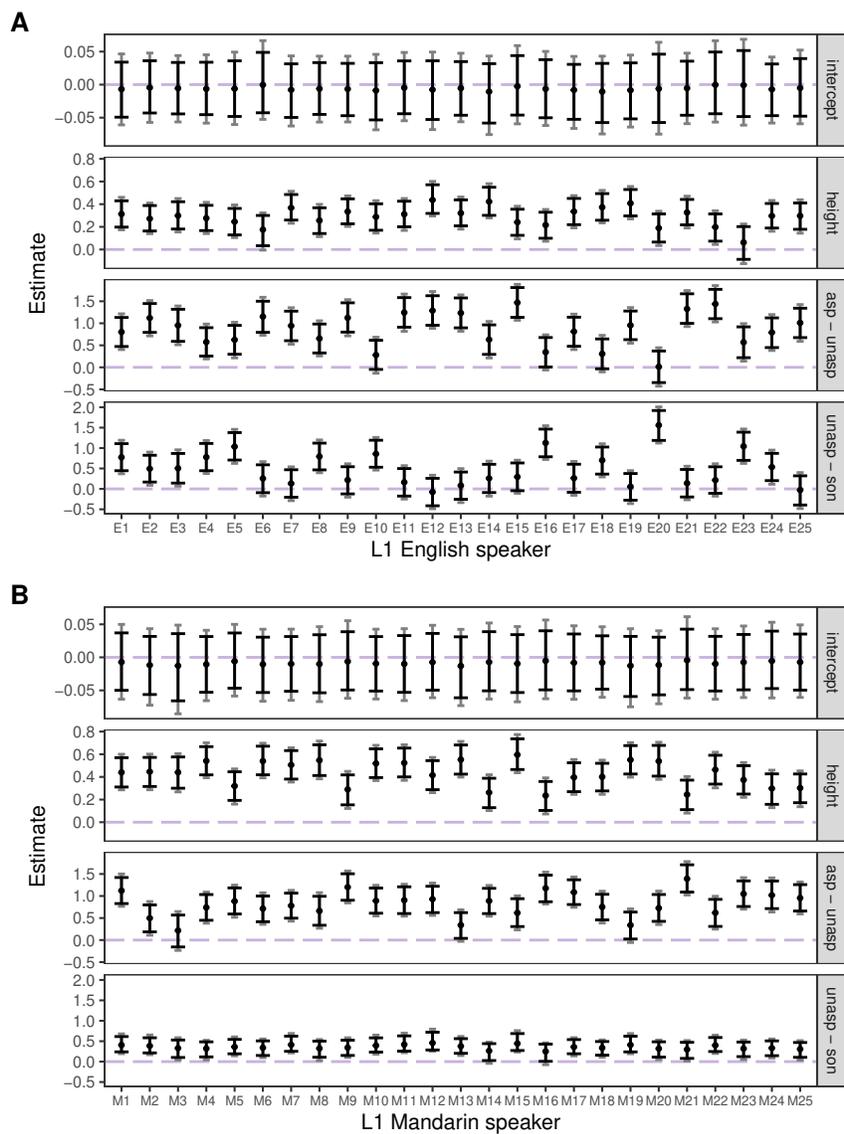


**Figure 4.3:** Population-level parameters from M2. Inner error bars represent 89% CrIs, and outer error bars represent 95% CrIs. The posterior mean for each parameter is indicated by a dot.

### Post-stop F0 Weights in L1 English and L1 Mandarin Speakers' English Production

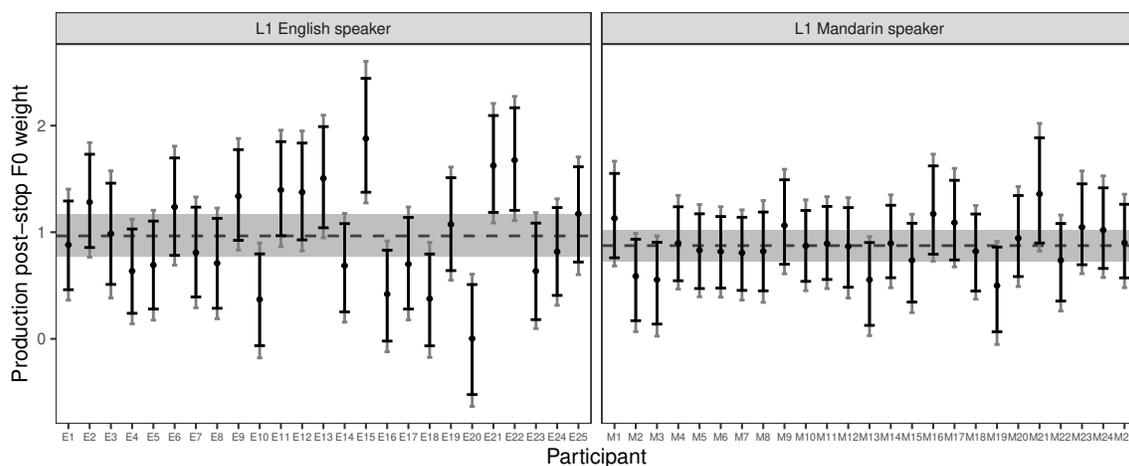
Figure 4.5 summarizes the post-stop F0 weights in production for the two speaker groups, at the population as well as individual levels. The detailed values the figure is based on are provided in Table E.4 in the appendix.

As can be seen in Figure 4.5, the two speaker groups had very similar production weights for post-stop F0 (L1 English: Mean = .97, 89% CrI = [.77, 1.16], L1 Mandarin: Mean = .87, 89% CrI = [.73, 1.02],  $\text{Diff}_{\text{L1 English} - \text{L1 Mandarin}}$ : Mean = .09, 89% CrI = [-.16, .33]), but more individual variation was present in the L1 English group. In particular, while the 89% CrI of every L1 Mandarin speaker overlapped with the 89% CrI of the population mean weight, there were four L1 English speakers (i.e., E15, E20, E21, and E22) whose post-stop F0 weights were



**Figure 4.4:** Marginal posterior summaries of parameters from M2 for participants in each language group. The dots represent the posterior means while the error bars enclose parameter values within 89% CrI.

either substantially higher or lower (i.e., there was no overlap between the 89% CrI of an individual speaker and the 89% CrI of the mean weights) than the population mean (note that speaker E20 even had a post-stop F0 weight of around 0).



**Figure 4.5:** Production post-stop F0 weights, as approximated by Cohen’s  $d$ , for L1 English and L1 Mandarin speakers. The posterior means of population-level weights are marked by the dashed line, and the shaded areas represent the 89% CrI of the mean weights. The posterior means of weights at the individual level are denoted by the dots, with individual error bars defining the 89% CrI.

#### 4.2.6 Interim Discussion: Production

Comparing L1 Mandarin and L1 English speakers’ production reveals that the two groups had a similar use pattern for post-stop F0. That is, the F0 after an aspirated stop was on average higher than that after an unaspirated stop, and the F0 after an unaspirated stop was in turn higher than that following a sonorant. The two groups also had similar mean production weights for post-stop F0 at the population level. At the individual level, however, there seemed to be more variation in the production weight for L1 English speakers than for L1 Mandarin speakers. More discussion pertaining to these findings will be presented in Section 4.4.2.

## 4.3 Perception Experiment

This experiment tapped into L1 English and L1 Mandarin listeners' use of post-stop F0 in perceiving stop voicing contrasts in word-initial position.

### 4.3.1 Participants

The same groups of participants from the production experiment also served as participants for the perception experiment, so the perceptual data analyzed here came from the same 50 people whose production data was analyzed. Note that the perceptual data from the L1 Mandarin listeners is a repeat from the English perceptual data from Chapter 3.

### 4.3.2 Stimuli and Procedure

The stimuli and procedure were exactly the same as the English version of the perception experiment described in Section 3.4 from the previous chapter.

### 4.3.3 Statistical Analyses

The analysis procedure followed the same basic steps as laid out in Section 3.4.6. Raw VOT and post-stop F0 values were standardized before being used in the model. The variable tone was sum-coded with TONE 1 and TONE 4 being coded with 1 and  $-1$  respectively. For the dependent variable, the /b/ response was coded with 0 while the /p/ response was coded with 1, so a positive coefficient means that higher values of the corresponding dimension elicit more /p/ responses in listeners than lower values.

The model employed was a hierarchical logistic mixture model. What differentiated the current model from the one used in the previous chapter was the random-effects/hierarchical structure behind the logistic regression. Since the experiment here involved a between-subject design, the random effects assumed a structure that was the same as the one used in the post-stop models described in Section 4.2.4. That is, each logistic parameter was decomposed into three parts: a language-universal estimate, a language-group-specific adjustment, and a participant-specific adjustment. Each language-universal estimate had as the prior a normal distribution with mean 0 and standard deviation 10. Language-group-specific adjustments

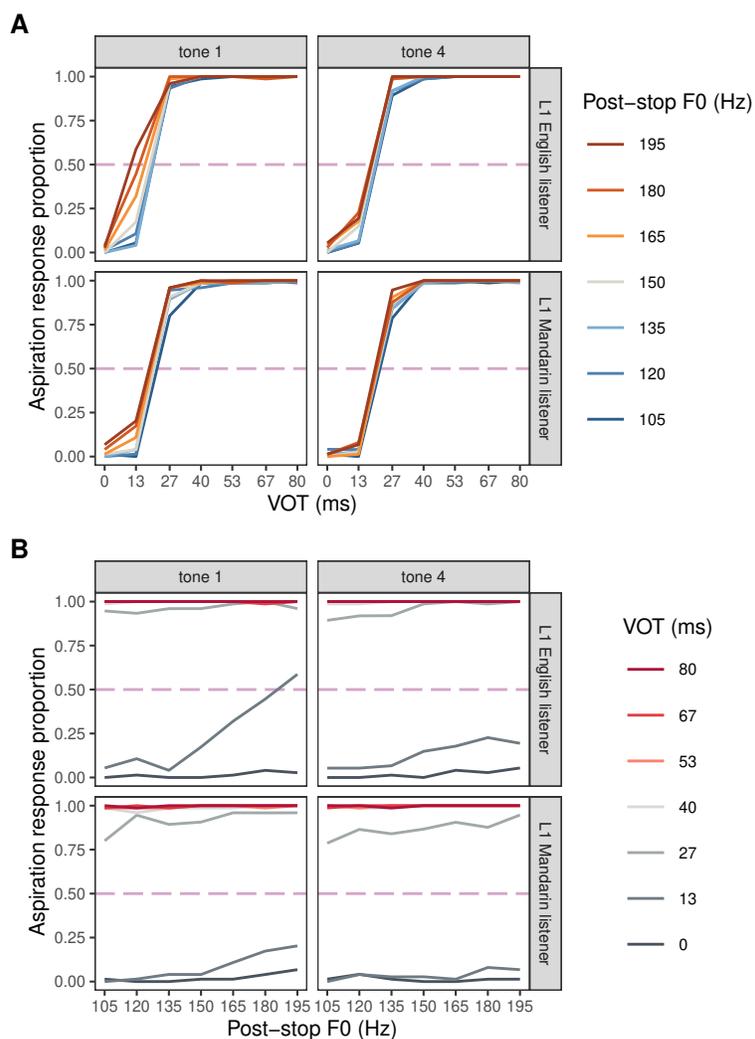
assumed a multivariate normal distribution for correlation, with the correlation matrix having an LKJ prior with  $\xi = 1$ . The correlations among participant-specific adjustments within each language group were similarly modeled using a multivariate normal distribution with an LKJ prior of  $\xi = 1$ . Finally, an exponential distribution with a rate of 1 was chosen as the prior for all variance parameters needed in the hierarchical model. The mathematical formulation of the model can be found in Section E.2.2 in the appendix.

In a similar vein to the previous chapter, the specifications of candidate models reflected prior knowledge and a balance between model complexity and predictive accuracy. **VOT** was automatically included in each model, as it is well-established to be the primary dimension L1 English and L1 Mandarin listeners rely on to distinguish aspirated and unaspirated stops (e.g., Guo, 2020; Schertz et al., 2020). The base model in this case included VOT as the only predictor, with first-order and second-order terms involving other factors—post-stop **F0** and **tone**—being added incrementally to form more complex models. The full set of models compared is given in Table 4.6.

#### 4.3.4 Results

The response patterns over different VOTs, post-stop F0s, and tones are displayed in Figure 4.6 for both listener groups. The ELPD-LOO means and standard errors for candidate models are provided in Table 4.6, and the model comparison results are summarized in Table 4.7.

Model comparison results spoke to the importance of post-stop F0 and tone in predicting listeners' responses (post-stop F0: M1 vs. M2 and M3 vs. M4; tone: M1 vs. M3 and M2 vs. M4). However, more complex models beyond M4 that had interaction terms between dimensions only improved model predictions marginally. As such, M4, which had just simple effects of **VOT**, **F0**, and **tone**, was selected as the final model for interpretation and discussion.



**Figure 4.6:** L1 English and L1 Mandarin listeners' aggregated categorization results of word-initial stops in English, shown as a function of VOT, post-stop F0, and tone. **A.** With VOT on the  $x$ -axis. **B.** With post-stop F0 on the  $x$ -axis to highlight its effect on categorization.

**Table 4.6:** Candidate perceptual models considered in model comparison, with their ELPD-LOO means and standard deviations.

<b>Model</b>	<b>ELPD-LOO</b>	<b>ELPD-LOO standard error</b>	<b>Predictors</b>
M1	-1392.7	51.5	VOT
M2	-1270.1	49.6	VOT + F0
M3	-1371.7	51.3	VOT + tone
M4 (final)	-1246.7	49.6	VOT + F0 + tone
M5	-1240.3	48.9	VOT + F0 + tone + F0 × VOT
M6	-1238.4	49.7	VOT + F0 + tone + F0 × tone
M7	-1236.0	49.2	VOT + F0 + tone + F0 × VOT + F0 × tone
M8	-1234.8	49.2	VOT + F0 + tone + F0 × VOT + F0 × tone + VOT × tone

**Table 4.7:** Model comparison results for key perception model pairs, expressed in differences in ELPD-LOO and associated standard errors. Pairs judged to differ in predictive power are marked by asterisks.

Model	M2	M3	M4	M5	M6	M7	M8
M1	-122.6* (15.4)	-21.0* (6.8)					
M2			-23.4* (7.5)				
M3			-125.0* (15.6)				
M4				-6.4 (5.9)	-8.3 (5.7)		
M5						-4.3 (5.0)	
M6						-2.3 (4.7)	
M7							-1.2 (2.3)

### Post-stop F0 Weights in L1 English and L1 Mandarin Listeners' English Perception: Population Results

The marginal posterior distributions for population-level effects from M4 are given in Table 4.8. All predictors had an effect on categorization, but the following descriptions focus particularly on two aspects: (1) Was post-stop F0 used as a cue for stop voicing by both listener groups? (2) If so, did the perceptual weights assigned to post-stop F0 differ by listener groups? With regard to the first question, the corresponding model parameters suggested a positive answer, with a higher post-stop F0 value leading to a higher probability of the token being classified as having an aspirated onset (L1 English:  $\bar{\beta} = 1.39$ , 89% CrI = [1.18, 1.62],  $p(\beta > 0) = 1.00$ ; L1 Mandarin:  $\bar{\beta} = 1.04$ , 89% CrI = [.76, 1.35],  $p(\beta > 0) = 1.00$ ). The second question can be answered by comparing the post-stop F0 weights of the two groups with each other, and the result indicated that there was weak evidence that L1 English listeners weighted the post-stop F0 cue more than L1 Mandarin

listeners ( $\bar{\beta} = .35$ , 89% CrI =  $[-.03, .67]$ ,  $p(\beta > 0) = .94$ ).

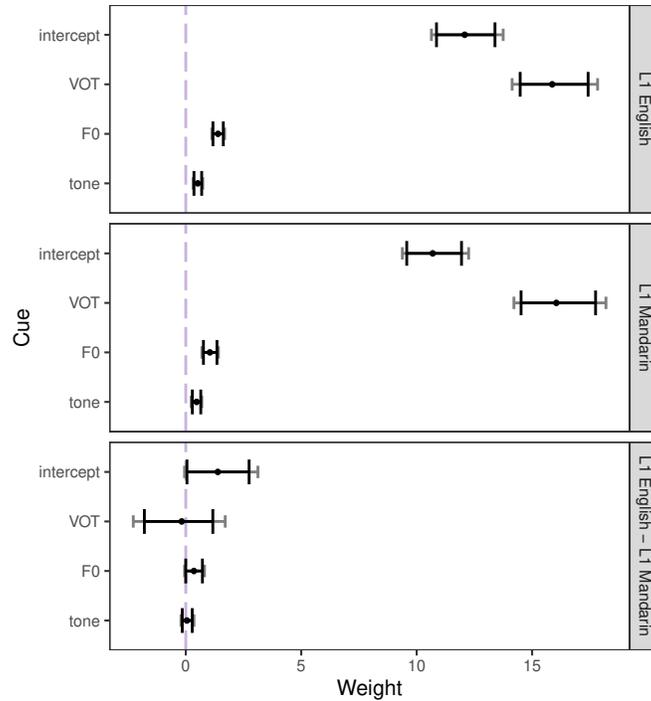
**Table 4.8:** Marginal posterior summary for key parameters from M4. The parameters whose effects are judged to be strong are marked with \*\*, and those whose effects are judged to be weak are marked with \*.

Parameter	Mean	SD	89% CrI	$p(\text{dir.})$
intercept <sub>ES</sub> **	12.08	.79	[10.86, 13.39]	$p(\beta > 0) = 1.00$
VOT <sub>ES</sub> **	15.87	.92	[14.48, 17.42]	$p(\beta > 0) = 1.00$
F0 <sub>ES</sub> **	1.39	.14	[1.18, 1.62]	$p(\beta > 0) = 1.00$
tone <sub>ES</sub> **	.52	.11	[.35, .69]	$p(\beta > 0) = 1.00$
intercept <sub>MS</sub> **	10.69	.74	[9.57, 11.94]	$p(\beta > 0) = 1.00$
VOT <sub>MS</sub> **	16.05	1.02	[14.52, 17.75]	$p(\beta > 0) = 1.00$
F0 <sub>MS</sub> **	1.04	.19	[.76, 1.35]	$p(\beta > 0) = 1.00$
tone <sub>MS</sub> **	.47	.12	[.28, .65]	$p(\beta > 0) = 1.00$
intercept <sub>ES</sub> – intercept <sub>MS</sub> *	1.38	.84	[-.05, 2.53]	$p(\beta > 0) = .96$
VOT <sub>ES</sub> – VOT <sub>MS</sub>	–.18	.90	[-1.85, 1.11]	$p(\beta < 0) = .58$
F0 <sub>ES</sub> – F0 <sub>MS</sub> *	.35	.22	[-.03, .67]	$p(\beta > 0) = .94$
tone <sub>ES</sub> – tone <sub>MS</sub>	.05	.14	[-.16, .27]	$p(\beta > 0) = .63$

### Post-stop F0 Weights in L1 English and L1 Mandarin Listeners' English Perception: Individual Results

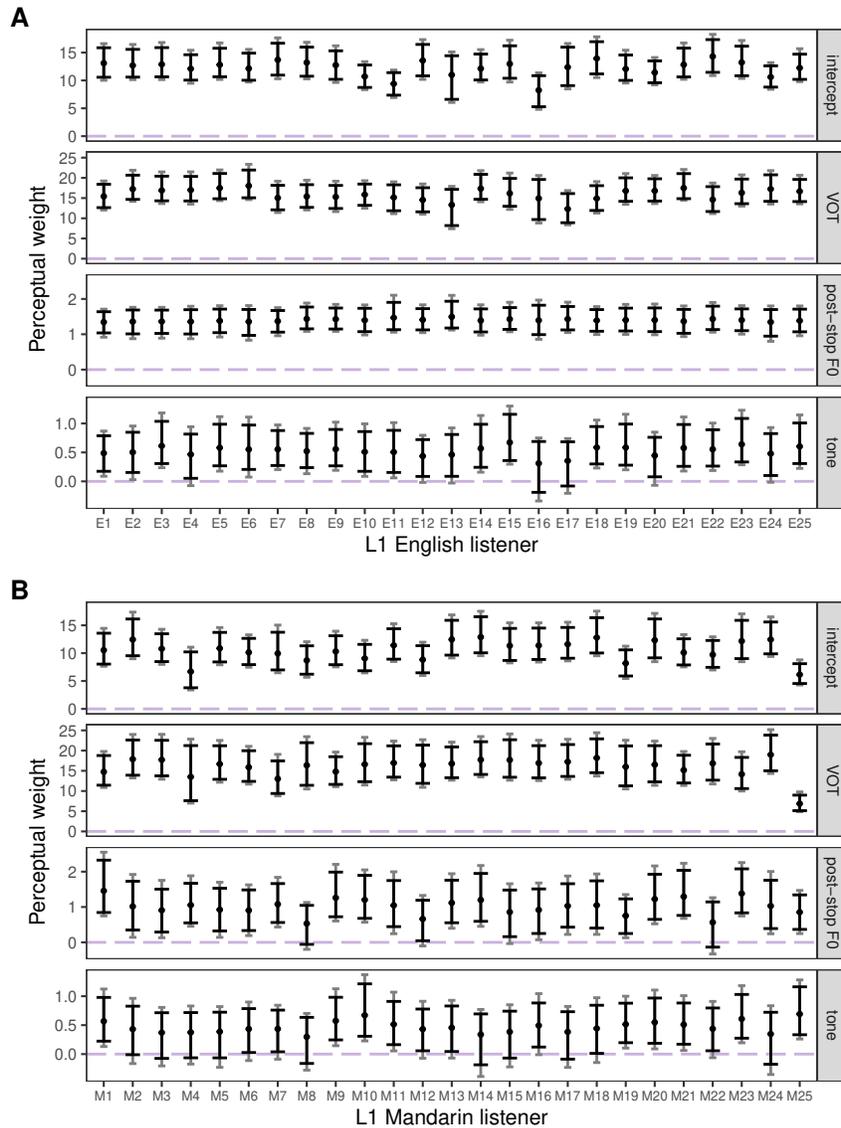
Shifting gears to post-stop weights at the individual level, Figure 4.8 shows the posterior distributions of individual post-stop F0 weights, along with the weights of other dimensions. Overall, all listeners, regardless of their L1, demonstrated a robust use of the post-stop F0 cue, in the same direction as the population tendency. Focusing on the post-stop F0 weights of L1 English listeners, this listener group was relatively homogeneous on the post-stop F0 cue, with all individuals having a mean weight around 1.4. In contrast, L1 Mandarin listeners were characterized by more variability—the mean weights ranged from .5 to 1.5—which is the opposite of the production pattern, where L1 Mandarin speakers had less variability. Additionally, there was more uncertainty surrounding each estimated weight.

Now with both population-level and individual-level results on the table, a richer picture is revealed. Even though there was only weak evidence that L1



**Figure 4.7:** Population-level parameters from M4. Inner error bars represent 89% CrIs, and outer error bars represent 95% CrIs. The posterior mean for each parameter is indicated by a dot.

English listeners relied on the post-stop F0 cue more than L1 Mandarin listeners at the population level, this might be caused by increased individual variation present in L1 Mandarin listeners. In fact, as far as the posterior means were concerned, individual L1 English listeners appeared to assign on average more weight to post-stop F0 than individual L1 Mandarin listeners. Therefore, it would be more accurate to conclude that in spite of L1 Mandarin listeners coming close to L1 English listeners with respect to the reliance on post-stop F0, the two groups were nonetheless not the same, with some L1 Mandarin listeners on a par with L1 English listeners and others making less use of post-stop F0 than L1 English listeners ( $\sigma_{L1\text{ Man.}}^2 - \sigma_{L1\text{ Eng.}}^2: \bar{\beta} = .18, 89\% \text{ CrI} = [-.08, .47], p(\beta > 0) = .87$ ).



**Figure 4.8:** Individuals' estimated weights from the perceptual model. **A.** Distributions of individual weights along various dimensions for L1 English listeners. **B.** Distributions of individual weights along various dimensions for L1 Mandarin listeners. Posterior means are represented by the dots, and the 89% CrIs are marked with error bars.

### 4.3.5 Further Analysis: Comparing Post-Stop F0 in Production and Perception

Both the SLM and PAM-L2 are based on a tight link between production and perception, so the two theories would predict that, in the present research context, individual production weights would be correlated with individual perception weights. The most rigorous way to verify this prediction involves fitting a model with both production and perception data, so that uncertainties within and across modalities can be properly handled. However, such a model necessarily requires a richer structure and computational power than the accessible hardware allows. As a surrogate model, individuals' post-stop F0 weights in production, represented by each person's *mean* Cohen's *d*, was correlated with their post-stop F0 weights in perception, approximated by the *mean* beta-coefficient from the logistic regression model. This approach therefore leaves out information about the uncertainties associated with both production and perceptual weights in the surrogate model.

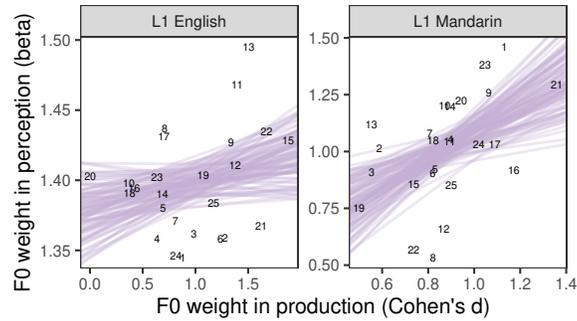
The individual data used in the correlational analysis and the resulting correlation lines are plotted in Figure 4.9. Both language groups showed a positive correlation trend, with L1 Mandarin speakers ( $\bar{\rho} = .45$ , 89% CrI = [.18, .68],  $p(\rho > 0) = .99$ ) demonstrating a stronger correlation than L1 English speakers ( $\bar{\rho} = .26$ , 89% CrI = [-.06, .53],  $p(\rho > 0) = .91$ ). However, as repeatedly emphasized, the correlation analysis did not take uncertainty around the mean weights into account, so better models and/or more data will be needed to draw a more definite conclusion.

## 4.4 Discussion

### 4.4.1 Summary of Results

This chapter compares L1 Mandarin-L2 English bilinguals' use of post-stop F0 in English production and perception with that of L1 English speakers. The connection between the results and the predictions presented in Section 4.1.4 is specified in Table 4.9 and discussed below.

In production, L1 Mandarin speakers' post-stop F0 patterns and magnitudes were similar to those observed for L1 English speakers at the population level: the



**Figure 4.9:** Post-stop F0 weights across perception and production for each L1 English and L1 Mandarin participant. Solid lines represent 100 regression lines fit with 100 posterior draws, to show direction and uncertainty in the correlation.

F0 following a voiceless stop tended to be higher than that after a voiced stop, and the F0 after a voiced stop was in turn higher than that following a sonorant. This trend was also reflected in both speaker groups' production weight for post-stop F0: at the population level, the mean production weights of both speaker groups were very similar. However, the two groups differed with respect to the amount of individual variation present, such that individual weights for L1 English speakers spanned a wider range than the individual weights for L1 Mandarin speakers.

Both speaker groups also used post-stop F0 as a cue for stop voicing in perception. Based on the 89% CrI, the model did not indicate strong evidence that the two listener groups relied on the post-stop F0 cue to different degrees at the population level. However, a careful inspection of individual weights reveals that there was greater between-listener variation in L1 Mandarin listeners, and that for all L1 Mandarin listeners, there was more uncertainty surrounding the estimated post-stop F0 weights, as compared to L1 English listeners.

Also, when individuals' mean production weights were correlated with the same individuals' mean perception weights, a weak positive correlation was observed for both language groups. However, as the performed correlation analyses did not incorporate the uncertainties inherent in the estimates, the apparent positive relationship across modalities was likely to be overestimated.

The following sections provide possible explanations and point out further is-

sues in relation to the mismatches between predicted and actual results, focusing on individual variation in L1 and L2 speech, the effect of L1 perceptual interference on L2 perception, and the production-perception interface.

**Table 4.9:** Predicted and actual results under different frameworks.

Framework	Predicted production	Predicted perception	Match experiment results?
SLM and PAM-L2	- Post-stop F0 weight: L1 English speakers > L1 Mandarin speakers - Perception should mirror production.		✗ - Mean post-stop F0 weight in production: L1 English speakers $\approx$ L1 Mandarin speakers - Mean post-stop F0 weight in perception: L1 English speakers > L1 Mandarin speakers (weak evidence) - Mismatch between production and perception
Perceptual interference	—	- Post-stop F0 for L1 Mandarin listeners has a lower weight and/or more variability	✓ - Mean post-stop F0 weight in perception: L1 English speakers > L1 Mandarin speakers (weak evidence) - L1 Mandarin listeners showed more between- and within-subject variations

#### 4.4.2 Individual Variation in Native and Non-Native Speech

Juxtaposing production and perception results, while at the population level, the two language groups seemed to behave similarly, results at the individual level appeared less consistent: whereas more individual variation was observed for L1 English speakers for production weights, it is L1 Mandarin listeners that manifested more individual variation when it comes to perception weights.

It might seem surprising that L1 English speakers showed more variation in their post-stop F0 production weights for their L1 English, given that prior work generally finds non-native speech to be more variable than native speech (e.g., Jongman and Wade, 2007; Wade et al., 2007). However, the bulk of these previous studies focuses on the L2 phonological contrasts that are absent in the speakers' L1. For instance, Wade et al. (2007) investigate the production of vowel spectral prop-

erties of L1 Spanish learners of English, as compared with L1 English speakers. They find greater group-level within-vowel category variability for the L2 speakers as compared with the L1 speakers, as measured by pooled standard deviation of F1 and F2. In the context of the current research, though the detailed phonetic realization of phonological voicing in stops differs by the language, the two-way stop voicing contrast in English poses no problem for L1 speakers of Mandarin, which also has a two-way contrast in stop voicing. L1 Mandarin speakers should therefore experience no difficulties in acquiring the English contrast. The observation that L2 speech tends to be more variable might thus not be applicable to the current study. In fact, recent studies (e.g., Vaughn et al., 2019; Xie and Jaeger, 2020) have reported that L2 speech is *not* necessarily more variable than the speech by native speakers of the L2, at least as far as individual-level within-speaker variability is concerned.

The account above still leaves open the question why there is more *between-speaker* L1 variation in English production in this study. Three explanations, though speculative in nature, are put forth below, both of which point to the influence of L1 Mandarin on L2 English. First, around half of the L1 English participants in this study can speak another (non-tonal) language reasonably well and therefore are not monolingual English speakers (see Table E.1 in the appendix). In fact, it is more difficult to find monolingual English speakers than multilinguals in Canada. The increased between-speaker variation in L1 English production might then be due to the heterogeneity of the L1 English group.

Second, it might be the case that L1 Mandarin speakers have a smaller individual variation in L2 English production because post-stop F0 is a dimension that does not manifest much individual variation (see Figure 3.18 from the previous chapter) even in Mandarin. The smaller individual variation simply reflects the carryover of an L1 characteristic to the production of L2. A simple way to test this account is to collect and compare more English production data from L1 Mandarin and L1 English speakers.

The third account, which is inspired by Xie and Jaeger (2020), relies on the premise from the SLM that phonetic categories in L1 and L2 operate in the same acoustic-phonetic space, and the argument based on the results from the Chapter 3 that L1 Mandarin speakers in this study have in fact established separate phonetic

categories for Mandarin and English stops. Now with more phonetic categories than Mandarin monolinguals, the acoustic-phonetic space of bilinguals becomes more crowded. In order to maintain distinctions between L1 and L2 categories, and between categories of the same language, the speaker needs to not only shift the centers of these categories away from one another, but also constrain the variability of these categories to reduce the overlap between the distributions of these categories in the acoustic-phonetic space. The reduced variability in L1 Mandarin speakers' production is therefore a byproduct of the strategy to maintain multilingual contrasts on the part of L1 Mandarin speakers. Note that this account does not preclude the first account, and one way to test it is by comparing the variability of the same phonetic category from bilingual and monolingual speakers of Mandarin (Kartushina and Frauenfelder, 2014).

#### **4.4.3 The Influence of L1 Perceptual Interference on L2 Stop Perception**

In contrast to the production results, the perception model indicates a rather homogeneous post-stop F0 weight distribution for L1 English listeners and a more spread weight distribution, in terms of within-listener and between-listener variability, for L1 Mandarin listeners.

This observation aligns well with the perceptual interference account (Lee and Nusbaum, 1993; Repp and Lin, 1990) put forth in Section 4.1.4. The fact that L1 Mandarin listeners process consonant and tone in a more integrated way, and that they are sensitive to tonal variation even when the tonal dimension is irrelevant for the task in question indicates that tonal information consistently competes for attention with variation along other acoustic dimensions. Given that post-stop F0 was manipulated in addition to the syllable tonal contour, post-stop F0 variation might be perceived as random fluctuations associated with tone. Effectively, this diminishes the intended perceptual magnitude of post-stop F0 changes, making it a less reliable cue for voicing. Numerically, the reduction in reliability causes the post-stop F0 weights to assume a wider range, for both within- and between-listener distributions. Of course whether this explanation is correct requires further research.

#### 4.4.4 Production-Perception Interface

The relationship between production and perception was analyzed by comparing each participant's mean production weight with their mean perception weight. As discussed above, with only individual *mean* weights, the correlational analyses suggested a positive trend between production and perception weights at the individual level. However, this positive trend is likely to be weakened when uncertainties surrounding the mean values are taken into account. This potential weaker association between production and perception if uncertainty is included calls into question the Motor Theory (Liberman and Mattingly, 1985) and the Direct-Realist approach's (Fowler, 1986) claim that there is a direct link between the two modalities in the L1, as well as the assumptions underlying the SLM and PAM-L2 that L2 perception preceded production. There are also a number of studies reporting that L2 learners can produce L2 sounds more accurately than perceive them (e.g., Flege et al., 1997; Flege and Eefting, 1987a,b; Trofimovich and John, 2011). For example, Trofimovich and John (2011) report that, while L1 Quebec French learners of English are able to produce /θ/ and /ð/ accurately, they have difficulties perceiving the contrast in a lexical decision task. In fact, the most recent iteration of the SLM—SLM-r—has acknowledged that even though the two modalities may coevolve (as opposed to perception leading production, where the accuracy of L2 segmental perception places an upper limit on the accuracy with which L2 sounds are produced, as proposed in the SLM), they might never align exactly (Flege and Bohn, 2021).

Methodological and analytical factors are also potential barrier to finding a relationship (Schertz and Clare, 2019). First, standard production and perception tasks are not entirely parallel, both in the nature of the experimental tasks and in the acoustic-phonetic space tapped into by the types of tasks. In particular, standard production tasks allow the speaker to produce cues in an unconstrained manner; in perception, however, the manipulated acoustic-phonetic space often extended to areas unlikely to be produced naturally. A lack of apparent link between the two modalities might therefore be caused by different acoustic-phonetic spaces being targeted by the two tasks. Second, the different natures of the production and perception tasks also have implications for statistical analyses that can be applied to

quantify weights. For instance, some cues are often strongly correlated in production, rendering production weights based on linear regression techniques unreliable due to collinearity. On the other hand, since cues are typically manipulated orthogonally in perception tasks, methods based on linear regression are commonly used to determine perceptual weights. Given that different metrics give different results, it is perhaps less surprising that whether and how production and perception are interdependent partially hinges on the chosen metrics. Lastly, the production and perceptual weights used in this study are technically different *kinds* of weights. The production weight is adopted from a type of distance-based measure, which quantifies the degree of separation between the two categories, while the perceptual weight is derived from the coefficient in a logistic regression, which has a probabilistic interpretation. To understand the production-perception interface therefore also involves a better understanding of how various kinds of metrics impact the interpretation of correlation.

## 4.5 Conclusion

The current work provides a matched set of perception-production data from 25 L1 English speakers and 25 L1 Mandarin-L2 English speakers, focusing on the relative role of post-stop F0 in distinguishing the stop voicing contrast in English. Whereas at the population level the two groups behaved similarly in both production and perception, cue weights at the individual level suggest a more nuanced picture. In particular, L1 English speakers' production weights showed more between-subject variation, but L1 Mandarin listeners' perceptual weights exhibit both more between-subject and within-subject variation. The asymmetric variational patterns are attributed to the potential heterogeneity in the language experience of L1 English group, to L1 Mandarin-L2 English bilinguals' L1 transfer of smaller individual variation, to the bilinguals' pressure to maintain distinct phonetic categories in response to their more packed common phonological space, and to a strong perceptual interference between consonant and tone for the bilinguals.

On the front of production-perception link, correlation analyses with mean weights showed positive trends for both groups, though the correlations are likely to be weaker when uncertainties are incorporated into the correlation models.

While developing a full-fledged model to fit production and perception data simultaneously and therefore to arrive at a more definite answer with regard to the strength of the production-perception interface is beyond the scope of this study, methodological and analytical decisions will also inevitably affect the findings.

## **Chapter 5**

# **The Effect of Cognitive Load on the Use of VOT and Post-Stop F0 in the Perception of Stop Voicing Contrasts by L1 English and L1 Mandarin Listeners**

### **5.1 Introduction**

Attention is a cognitive process that selectively filters information. Selective attention also plays a role in speech processing because listening conditions in everyday life often demand the listener to selectively attend to linguistic signals affected by factors such as environmental noise, impaired hearing ability, second language communication, and sound change in progress. In lab settings, adverse listening conditions can be induced by manipulating cognitive load (CL) while the listener performs speech perception tasks (Gordon et al., 1993; Kong and Lee, 2018; Mattys et al., 2014; Mattys and Palmer, 2015; Mattys and Wiget, 2011; Mitterer and Mattys, 2017). This type of manipulation is meant to simulate situations where the listener recognizes speech while performing another task. CL is defined as

“any load whose effect on speech recognition arises not from an energetic distortion of the signal but from the recruitment of central processing resources due to concurrent attentional or mnemonic processing” (Mattys and Wiget, 2011, p. 145). Behind this definition is the core assumption that attention and short-term memory are resources in limited supply, and that divided attention or short-term memory overload taps into these resources needed for speech processing. In connection to cue-weighting in speech perception, it has been reported that listeners attend to the primary cue in optimal listening conditions, while secondary cues become more prominent in adverse conditions, such as those induced by additional CL (Gordon et al., 1993; Kong and Lee, 2018). This suggests that listeners adapt the weights associated with various cues according to the ambient communicative context, and cue weights are modulated by the availability of processing resources.

This study revisits the relationship between cue weighting and selective attention, focusing on whether and how the use of voice onset time (VOT), the primary cue, and post-stop fundamental frequency (F0), a secondary cue, is affected by cognitive load in listeners who speak English or Mandarin as their first language (L1). In what follows I first review studies that address similar topics and whose methods inspire the approach of the current study. I then spell out the research questions and lay out the expected results derived from the findings of the reviewed studies. Detail about the conducted experiments is provided next, with a discussion on the results ensuing.

### **5.1.1 Attentional Modulation of Cue Weight**

An early but important work that probes the interaction between selective attention and cue weighting is Gordon et al. (1993). In one of their experiments, they investigated whether the amount of attention that was allocated to speech perception influenced the relative importance of two acoustic cues—VOT (i.e., the primary cue) and F0 (i.e., a secondary cue)—to the voicing distinction between the consonants /b/ and /p/. The experiment involved a within-subject design, where 18 L1 English listeners were presented with 10 experimental blocks alternating between the CL and non-CL conditions. The audio stimuli in both conditions were drawn from a /ba-/pa/ continuum, where the VOT of the initial consonant and the F0 of the

entire vowel were manipulated. VOT was manipulated in seven equal-increment steps from 10 ms to 65 ms, while the F0 of the vowel was fixed at either 100 Hz or 180 Hz. Note that the manner in which F0 was manipulated in their experiment differed in an important way from the post-stop F0 manipulation used in the current study: while only the initial 35% of F0 contour was altered in the current study (see Section 3.4.2 for further detail), the entire F0 contour was given a constant F0 in their experiment. The additional cognitive load was induced by having listeners perform a visually presented arithmetic distractor task (specifically, listeners were given three two-digit numbers which were all multiples of 10 and were asked to decide whether the difference between the first and second numbers was the same as the difference between the second and third numbers). In the CL blocks, a speech stimulus was presented 150 ms after the appearance of numbers for the arithmetic task, and the listener had to make a response in the arithmetic task before being prompted to identify the sound as /b/ or /p/ in a two-alternative forced choice (2AFC) task. In the non-CL blocks, three pairs of zeros appeared as the speech sound was presented, and the listener simply did the same 2AFC task. The analyses revealed a significant interaction between CL condition and the effect of VOT and F0. That is, when CL was increased, the listeners' reliance on the primary VOT cue was mitigated, but reliance on the secondary F0 cue was boosted, though the order of the weighting between these two cues remained unchanged. Gordon et al. (1993) interpreted their results as supporting the claim that there are differential degrees of attention in processing hierarchical linguistic information where higher order cues (e.g., VOT in the case of stop voicing in English) require greater attention in speech processing. More reliance on a lower-order cue (e.g., F0 in the case of stop voicing in English) under the CL condition might result from the reduced competition from the primary cue.

Using the same CL task and experimental paradigm as Gordon et al. (1993), Kong and Lee (2018) investigated the influence of attention on acoustic cue weightings in speech perception by examining 28 L1 Seoul Korean listeners' identifications of the three-way laryngeal stops in Korean (i.e., aspirated vs. lax vs. tense). The Seoul dialect has been found to undergo a sound change, and shift to use both VOT and F0 as primary cues for this three-way contrast in stop voicing (Kang and Guion, 2008; Kang, 2014; Kim, 2013; Lee and Jongman, 2012; Silva, 2006). That

is, VOT functions as the primary cue that singles out tense stops from aspirated and lax counterparts, and F0 serves as the primary cue to distinguish between aspirated and lax stops. Given this observation, the expectation was that CL should reduce the reliance on both cues. Their speech stimuli were prepared by manipulating VOT and F0 orthogonally along a /ta/-/t<sup>h</sup>a/ continuum. A VOT continuum with seven log-scale steps (9 ms, 13 ms, 19 ms, 28 ms, 40 ms, 58 ms, 100 ms) was created via cross-splicing. For each VOT step, a five-step F0 continuum was created by lowering and raising F0 from 98 Hz to 130 Hz in 8-Hz steps. Parallel to Gordon et al. (1993), each step of the F0 continuum had a consistent F0 for the entire vocalic portion. They found that VOT was an informative cue across the three stop laryngeal categories and the listeners' reliance on VOT was consistently weakened under the CL condition. The F0 cue, on the other hand, did not systematically interact with CL. Their results therefore agreed with the relationship observed in English stops, where a primary cue is associated with reduced perceptual dependency under limited attention. However, the fact that CL did not decrease the dependency on F0 consistently was at odds with the findings that Seoul Korean listeners use *both* VOT and F0 as primary cues for the three-way stop contrast. Because CL did not boost the perceptual reliance on F0, it was also not possible to define F0 as a secondary cue in perception, as in Gordon et al. (1993). The authors interpreted the lack of an enhanced role of F0 as indicating that the sound change in the stop laryngeal contrast has not been stabilized in Seoul Korean, and that listeners need to maintain flexible perceptual strategies to accommodate a co-existence of conservative (i.e., VOT as a primary information) and innovative forms (i.e., enhanced F0 information) of the stop categories (Beddor, 2009, 2015; Francis et al., 2008).

Before we conclude that CL serves as a proper diagnostic tool for assessing the relative importance of acoustic cues to phonological contrasts, however, it is crucial to point out that the nature of the distraction task also seems to play a critical part. For instance, as part of their research to investigate the interaction between the Ganong effect (Ganong, 1980) and CL, Mattys and Wiget (2011) tested speech categorization under non-CL and CL conditions in a between-subject design using a 2AFC task. In the experiment, L1 English listeners had to identify the initial stop of audio stimuli as /k/ or /g/. The audio stimuli were sampled from three eight-step continua, with VOT values ranging from 15 ms to 56 ms: *giss-kiss*

(nonword-word), *gift-kift* (word-nonword), and *gi* [gi]-*ki* [ki] (nonword-nonword). Different from Gordon et al. (1993) and Kong and Lee (2018), however, additional CL was created by having listeners perform a concurrent visual search task in an array of color shapes (see Section 5.2.2 for more detail). They did *not* find an effect of CL on the steepness of the categorization function, which is often viewed as reflecting the acuity of speech perception, at either the population or individual level. Several factors might contribute to the apparently contradicting results between Gordon et al. (1993) and Mattys and Wiget (2011). For instance, while Gordon et al.'s (1993) experiment followed a within-subject design, that in Mattys and Wiget (2011) used a between-subject design. In addition, the manipulation in Gordon et al. (1993) involved two acoustic dimensions: VOT and overall F0, but only VOT was manipulated in Mattys and Wiget (2011). The secondary task in Gordon et al. (1993) was sequential, but that in Mattys and Wiget (2011) was concurrent. In terms of the nature of the CL task, Gordon et al. (1993) used an arithmetic task as the secondary task whereas Mattys and Wiget (2011) adopted a visual search task. It is still unclear which combinations of these factors will induce which kinds of results. Last but not least, it might simply be that the experiment in Gordon et al. (1993) does not replicate. In fact, the current state of research features a broad array and heterogeneity of secondary tasks that lack standardization and continuity in their implementation, which makes cross-study generalizations difficult if not impossible (Bijarsari, 2021).

### **5.1.2 Goals of the Current Study**

This study examines L1 English and L1 Mandarin listeners' perception of their native stop-voicing contrasts under non-CL and CL conditions, with the experimental paradigm largely following that of Mattys and Wiget (2011). In particular, this case study focuses on whether and how CL modulates the perceptual weights of the primary VOT cue and a secondary post-stop F0 cue. The CL task of choice was a concurrent visual search task modeled after Mattys and Wiget (2011). This task was chosen because of its non-linguistic nature and non-auditory modality, so any effect on the speech perception is more likely to arise from depletion of central processing resources. In addition, a concurrent, as opposed to sequential, task was

used because it is more similar to everyday communication settings where signals from multiple modalities arrive at the same time.

Three participant groups—two groups of L1 English listeners and one group of L1 Mandarin listeners—participated in the CL perception experiment described in the next section. Together with the perceptual data from Chapter 3 and Chapter 4, three comparisons enable us to evaluate the effect of CL in the two languages (see Section 5.2.4 for more detail about the planned comparisons). The first is a within-subject CL versus non-CL comparison among a group of L1 English listeners, which makes this comparison closest to the within-subject design utilized in Gordon et al. (1993) and Kong and Lee (2018). The other two involve between-subject comparisons in English and Mandarin respectively, which makes them similar to the between-subject design in Mattys and Wiget (2011). Both within-subject and between-subject comparisons are used because they offer different advantages. While a within-subject design facilitates cross-condition comparisons at the individual level, a between-subject comparison has the advantage of mitigating practice effects in a within-subject design.

The expected results for the three comparisons from above are summarized in Table 5.1 and elaborated on below.

First, recall that the current experimental design differs from that of Gordon et al. (1993) or Kong and Lee (2018) in two important ways: (i) a visual search distractor task was used here instead of an arithmetic task, and (ii) the F0 manipulation here involved on the initial 35% of the vowel instead of the entire vocalic part. If these two places of divergence are of little importance (i.e., the visual search and the arithmetic task have a similar CL effect, and the two ways of F0 manipulation have a comparable perceptual consequence), then we would expect the results to follow those reported in Gordon et al. (1993). Note that the way in which the stimuli were manipulated in this study also differs from that of Mattys and Wiget (2011) in that both VOT and post-stop F0 were manipulated here whereas only VOT was modified in Mattys and Wiget (2011). It is possible that their finding that the weight did not change might not generalize to the current setting where two dimensions were manipulated at the same time. In addition, given the non-linguistic and non-auditory nature of the CL induced by the visual search task, it is expected that CL should exert similar impacts cross-linguistically. If the effect associated

**Table 5.1:** Expected results under different hypotheses.

<b>Hypothesis</b>	<b>Expected results</b>
The visual search task and the arithmetic task tap into the same central processing resources, and the post-stop F0 manipulation in this study has a similar effect as the overall F0 manipulation in Gordon et al. (1993).	VOT weight in CL ↓, and post-stop F0 weight in CL ↑ for all three comparisons.
The visual search task has no effect on cue weighting, as hinted on in Mattys and Wiget (2011), or the post-stop F0 manipulation in this study has a different effect than the F0 manipulation in Gordon et al. (1993). Alternatively, it could simply be that the findings in Gordon et al. (1993) are not replicable.	Weight in CL $\approx$ weight in non-CL for VOT and post-stop F0 in all three comparisons.

with CL is stable, we should also expect the general patterns in the within-subject comparison to mirror that in the subject-subject comparison, though the effect size might be smaller in the within-subject comparison due to practice effects. Overall, then, we would expect the VOT weight to decrease in the CL condition but the post-stop F0 weight to increase in the CL condition, regardless of the language.

Alternatively, the visual search task might tap into the central processing resources differently than does the arithmetic task. The different ways F0 was manipulated between the current work and Gordon et al. (1993) might also affect how the manipulated F0 was perceived. These two different design choices mean that what was observed in Gordon et al. (1993) might not be replicated here. Of course there is also the possibility that the results in Gordon et al. (1993) simply do not replicate. In fact, if the finding of Mattys and Wiget (2011) that CL did not modulate the use of the VOT cue can be generalized to the use of cues of other acoustic dimensions, we would expect CL to *not* have an effect on the VOT or post-stop F0 weight.

## 5.2 Perception Experiment with Cognitive Load

This experiment examined the effect of cognitive load on the perceptual cue use of VOT and post-stop F0 in L1 English and L1 Mandarin listeners. Since this experiment involved two listener groups, the experiment also came in two different versions: an English version for L1 English listeners and a Mandarin version for L1 Mandarin listeners. Crucially, both the English and Mandarin versions of the experiment followed the same design as the respective perceptual experiments in Section 3.4 and Section 4.3; the only difference was the addition of CL.

### 5.2.1 Participants

The participants included two groups of L1 English listeners and one group of L1 Mandarin listeners. The inclusion criteria for participants were the same as those laid out in Section 3.3.1 and Section 4.2.1 for Mandarin and English participants respectively.

#### L1 English Listeners

Two groups of L1 English listeners participated in the experiment. One group was the same L1 English speakers who completed the non-CL English production and perception experiments described in Section 4.2 and Section 4.3. In addition, this group of participants always did the non-CL experiments before the CL English perception experiment presented in this chapter. The other group consisted of 25 additional L1 English listeners (2 non-binary, 18 female, 5 male;  $\text{mean}_{\text{age}} = 20.6$  years,  $\text{median}_{\text{age}} = 20$  years,  $\text{SD}_{\text{age}} = 3.4$  years,  $\text{range}_{\text{age}} = [17, 33]$ ), who did not participate in any other previously described experiments. The demographic information of the additional L1 English listeners is given in Table F.1 in the appendix.

#### L1 Mandarin Listeners

The L1 Mandarin listener group included additional 25 participants that did not participate in the experiments described in Chapter 3 or Chapter 4 (21 female, 4 male;  $\text{mean}_{\text{age}} = 20$  years,  $\text{median}_{\text{age}} = 20$  years,  $\text{SD}_{\text{age}} = 1.9$  years,  $\text{range}_{\text{age}} = [18, 26]$ ). The demographic information of the 25 L1 Mandarin listeners is provided in Table F.2 in the appendix.

## 5.2.2 Stimuli

Additional cognitive load was induced by a concurrent visual search task. As such, apart from audio stimuli, this section also describes the visual stimuli used in the visual search task.

### Audio Stimuli

Depending on the version of the experiment, the audio stimuli were either the same as those used in the Mandarin perception experiment described in Section 3.4.2 or the same as those used in the English perception experiment described in Section 3.4.2. However, regardless of the version, the target stimuli were shared across the board. That is, only the fillers differed across the two versions, with those in the English version being [mi<sup>1</sup>], [mi<sup>4</sup>] (similar to English *me*), [ni<sup>1</sup>], and [ni<sup>4</sup>] (similar to English *knee*), and those in the Mandarin version being [i<sup>1</sup>] (*yi* 衣), [i<sup>4</sup>] (*yi* 意), [mi<sup>1</sup>] (*mi* 咪), and [mi<sup>4</sup>] (*mi* 密).

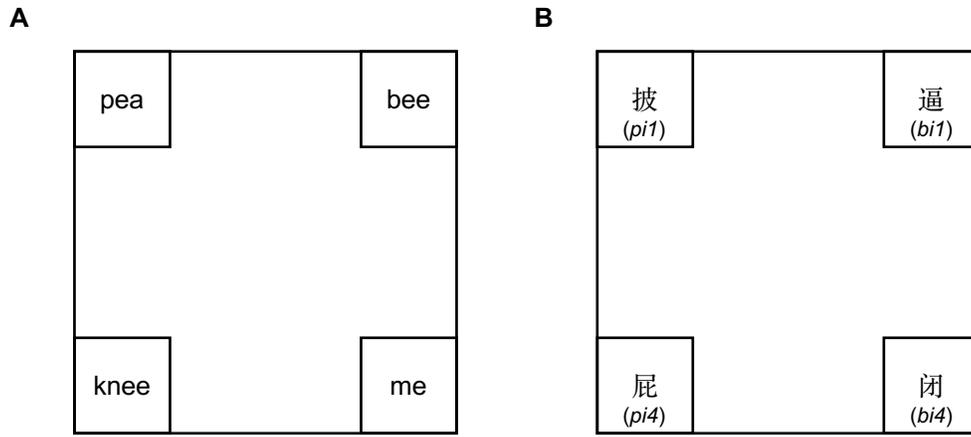
### Visual Stimuli

The visual distractors used in this experiment were modeled after those in Mattys and Wiget (2011). The distractors consisted of 500 grids made of six rows and six columns, with the size being approximately 9 cm by 9 cm when displayed on the screen. The 36 items in each cell were black squares or red triangles, arranged randomly in the grid. Half (= 250) of the grids contained a red square in a random cell, which is the target the participant was required to detect. One example from each of these two types of grids is shown in Figure 5.2.

## 5.2.3 Procedure

### Trial Response Configuration

The configuration of trials depended on the listener group. For the L1 English listener group that had also participated in the English production-perception experiment described in Chapter 4, the layout of the four response options was a within-subject factor, so, for instance, if a participant saw the layout in Figure 3.12A, repeated in Figure 5.1, for the non-CL perception experiment, the same layout would



**Figure 5.1:** A possible response layout for the English non-CL and CL experiments.

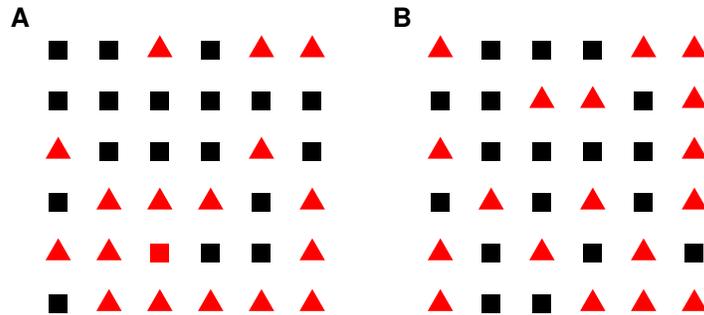
be used for this CL experiment. The rationale behind matching the layout across the non-CL and CL perception experiments was to control for variability when the results from the two experiments were compared.

For the other L1 English and the L1 Mandarin listener groups that were independently recruited solely for this experiment, the response layouts were counter-balanced within each listener group across all participating listeners, irrespective of whether their data was included in the analyses. That is, the distribution of the different layouts from the participants whose data was retained for the analyses might not be equally distributed.

### Experiment Procedure

The procedure largely followed that of the non-CL version of the perception experiment in the respective language (see Section 3.4.3 for the detailed procedure). The only differences were that participants in the CL experiment were additionally asked to pay attention to the distractor grid displayed on the screen during the playback of the audio stimulus and search for a red square, and to answer a question about the presence of a red square at the end of each trial.

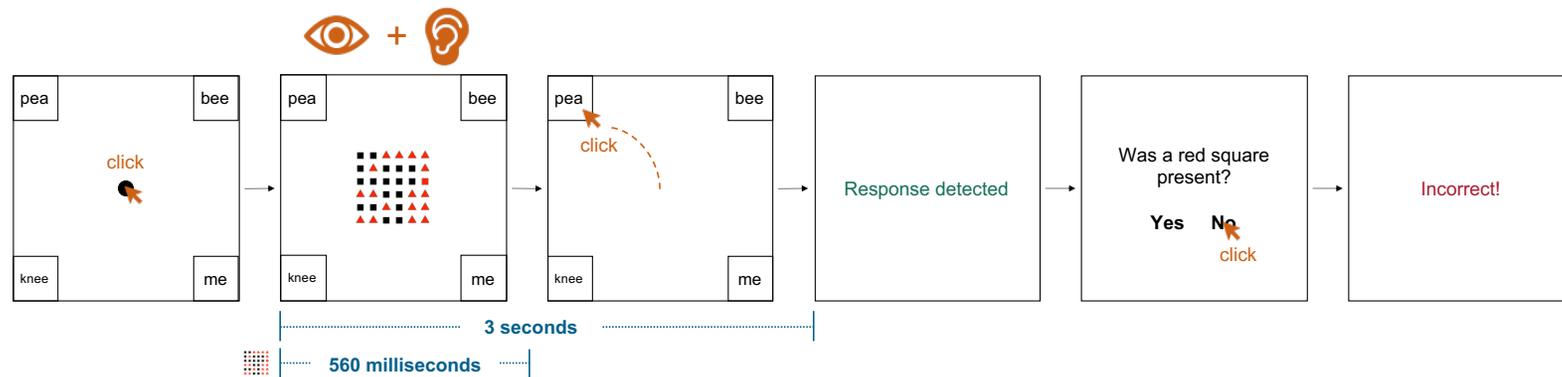
The timing of the elements within each trial is summarized in Figure 5.3 and described below. Upon the participant's clicking the center dot, the distractor grid



**Figure 5.2:** Examples of visual displays used as cognitive load. **A.** Target-present grid, with the target (i.e., red square) in the fifth row and third column. **B.** Target-absent grid.

was presented concurrently with the audio stimulus. To keep the level of CL constant across all trials, the grid was displayed for a fixed duration of 560 ms in all trials, starting at the same time as the audio stimulus (which has a duration between 350 ms and 430 ms). This duration was set based on the value used in Mattys and Wiget (2011).<sup>1</sup> Like the non-CL perception experiment, the participant had three seconds to respond by clicking on a word, timed from the onset of audio stimulus presentation. After clicking on a word, or at the end of a 3-second period, an additional question appeared on the screen, reading: “Was a red square?”. The participant had up to ten seconds to click one of two buttons—Yes or No—on the screen. They were then provided with feedback on whether their answer was correct with respect to the visual search task (but never to the audio four-alternative forced-choice task) and how many trials they managed to answer correctly in succession.

<sup>1</sup>With this particular combination of grid size and presentation duration, Mattys and Wiget (2011) report a mean accuracy of around 80%.



**Figure 5.3:** Illustration of trial procedure in the English cognitive load perception experiment. The overall procedure is similar to that of other perception experiments, except that the participant was additionally asked to pay attention to the grid during the playback of the syllable and search for a red square. The grid was displayed for 560 ms in all trials, and the participant had to indicate if they found a red square after they gave a response based on the audio stimulus.

#### 5.2.4 Statistical Analyses

The statistical models varied, contingent on the listener groups involved in the comparison. Three comparisons were carried out. The first was a within-subject comparison that contrasted the same L1 English listeners' performances across non-CL and CL experiments, with the listeners completing the non-CL experiment before the CL one. The second was a between-subject comparison that evaluated the performance of a group of L1 English listeners in the non-CL experiment on the one hand, and that of another group of L1 English listeners in the CL experiment on the other. It is worth emphasizing that performances from the two groups were indeed comparable because the tasks were the *first* perception tasks completed by both groups. The third was similar to the second, except that it involved the data from two groups of L1 Mandarin listeners in the non-CL and CL experiments respectively. As in the previous two chapters, only target trials with a "correct" response were included in the analyses. That is, only the target trials for which the listener selected *pea* or *bee* (or, for the Mandarin version, *pi1* or *bi1* for Tone 1 trials, and *pi4* or *bi4* for Tone 4 trials) as the response were considered.

The following sections spell out the sources of data and the statistical model used in each comparison.

##### **English Within-Subject non-CL versus CL Comparison**

The data consisted of L1 English listeners' data from the non-CL perception experiment described in Section 4.3 and the same listeners' perceptual data from the CL version just described.

As the analysis involved a within-subject comparison between a listener's performance in the non-CL experiment and that in the CL version, the model structure and the transformation and contrast coding of the predictor variables were identical to those used for comparing L1 Mandarin listeners' Mandarin and English response patterns, with one submodel predicting L1 English listeners' responses in the non-CL condition and the other submodel predicting the same listeners' responses in the CL condition. Details about the model and information on the priors can be found in Section 3.4.6.

### **English and Mandarin Between-Subject non-CL versus CL Comparisons**

For English, the data comprised the perceptual responses from the English non-CL perception experiment described in Section 4.3, and the perceptual responses from the additional 25 L1 English listeners participating in the English CL perception experiment just described. For Mandarin, the data came from the Mandarin non-CL perception experiment described in Section 3.4, which included the responses from 25 L1 Mandarin listeners, as well as from the Mandarin CL perception described above, which included the responses from additional 25 L1 Mandarin listeners.

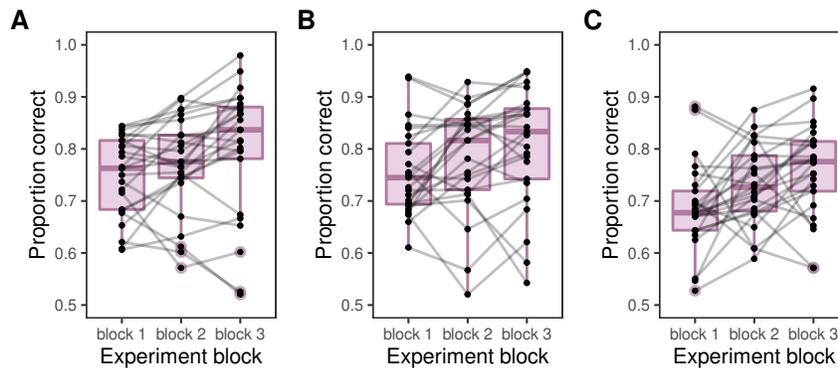
For both languages, the core of the analysis was a between-subject comparison with non-CL and CL perceptual data. The model structure and the transformation of the predictor variables were therefore the same as those specified for the comparison between L1 Mandarin listeners' and L1 English listeners' responses in the non-CL English perception experiment described in Section 3.4. Briefly, the model contained two submodels, with one submodel predicting the responses from the non-CL participants, and the other submodel predicting the responses from the CL participants. These two submodels were connected with a condition-agnostic covariance matrix. The detailed information about the formulation of the model and the prior is provided in Section 4.3.3.

### **5.2.5 Results: Visual Search Task**

The accuracy in the visual search task across the three blocks, separated for the three listener groups, is shown in Figure 5.4. Note that the accuracy was calculated based on only the target trials that the listener answered “correctly”, that is, those target trials with one of the target words being selected for the English version, or with one of the target words with the right lexical tone being selected for the Mandarin version. For completeness, the accuracy based on both “correct” target and filler trials is plotted in Figure F.1 in the appendix. Both sets of results are very similar to each other.

Since the accuracy in this task is not at the core of this study, no statistical analysis was performed on this set of data; suffice it here to show only the individual accuracy scores and to describe the trend. As can be seen in Figure 5.4, most listeners, regardless of the language group, had an accuracy score well above

.5 (chance) in all three blocks. In general, their accuracy also improved over the blocks. At the population level, the median fell at around .75 in the first block and climbed to about .81 in the third block for both L1 English groups, while the L1 Mandarin group had a slightly lower accuracy: from .68 in the first block to .79 in the third block. These numbers agree well with what has been reported in Mattys and Wiget (2011): their accuracy ranges between .6 and .9, with the mean at .8.



**Figure 5.4:** Accuracy in the visual search task, based on “correct” target trials (see Section 5.2.5 for detail). A score was estimated for each listener in each block. The scores from the same listeners are connected by a line. **A.** Results for L1 English listeners who also participated in the non-CL version. **B.** Results for L1 English listeners who only did the CL version. **C.** Results for L1 Mandarin listeners.

### 5.2.6 Results: Perceptual Weights

In order to make the model outputs from the three planned comparisons comparable to one another, the models in the three comparisons had the same fixed-effects predictor terms. The model comparison results from the English between-subject comparison were used to determine which terms were to be included in all models, given that this between-subject comparison was closest to the between-subject experimental design in Mattys and Wiget (2011). For this particular comparison, the candidate models considered are listed in Table 5.2, along with their ELPD-LOO means and standard errors. Table 5.3 summarizes model comparison results. The comparison between M1 and M2, and between M3 and M4, pointed to the

importance of including post-stop F0 as a predictor in the model. Similarly, the comparison between M1 and M3, and between M2 and M4, demonstrated the predictive power of tone as a predictor. However, models more complex than M4 with additional interaction terms did not seem to further boost the model's predictive ability. As such, M4 was selected as the final model, the predictor terms of which were included in all subsequent models.

**Table 5.2:** Candidate models considered in the model comparison for the English between-subject CL experiment, with their ELPD-LOO means and standard errors.

Model	ELPD-LOO	ELPD-LOO standard error	Predictors
M1	-1448.4	49.7	VOT
M2	-1251.5	47.5	VOT + F0
M3	-1426.1	49.5	VOT + tone
M4 (final)	-1220.9	47.6	VOT + F0 + tone
M5	-1221.2	47.6	VOT + F0 + tone + F0 × VOT
M6	-1216.3	47.7	VOT + F0 + tone + F0 × tone
M7	-1216.2	47.7	VOT + F0 + tone + F0 × VOT + F0 × tone
M8	-1219.3	47.9	VOT + F0 + tone + F0 × VOT + F0 × tone + VOT × tone

In presenting the results from individual comparisons in the following subsections, those from the English within-subject and between-subject comparison are given together since they showed the same patterns. The results from the Mandarin between-subject comparison are presented separately since their patterns deviated from those observed for English. As the focus of this study is on the potential changes in VOT and post-stop F0 weights in response to CL, the discussion will center around the results of these two cues. Following the convention from the previous chapters, findings at the population level are given before those at the individual level.

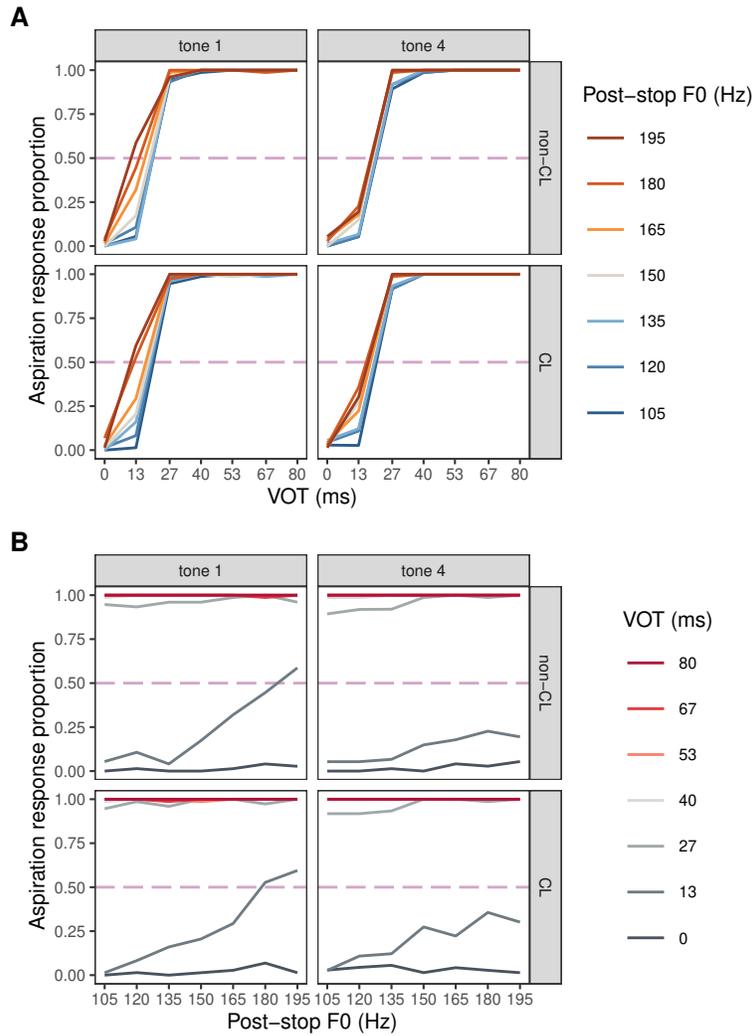
**Table 5.3:** Model comparison results for the English within-subject CL experiment. The numbers show differences in ELPD-LOO and associated standard errors for key model pairs. Pairs judged to differ in predictive power are marked by asterisks.

Model	M2	M3	M4	M5	M6	M7	M8
M1	-197.0* (18.7)	-22.3* (6.9)					
M2			-30.6* (8.2)				
M3			-205.3* (19.2)				
M4				0.4 (2.2)	-4.6 (4.6)		
M5						-5.1 (4.4)	
M6						-0.1 (1.7)	
M7							3.2 (0.9)

### English Within-Subject and Between-Subject Non-CL versus CL Comparisons

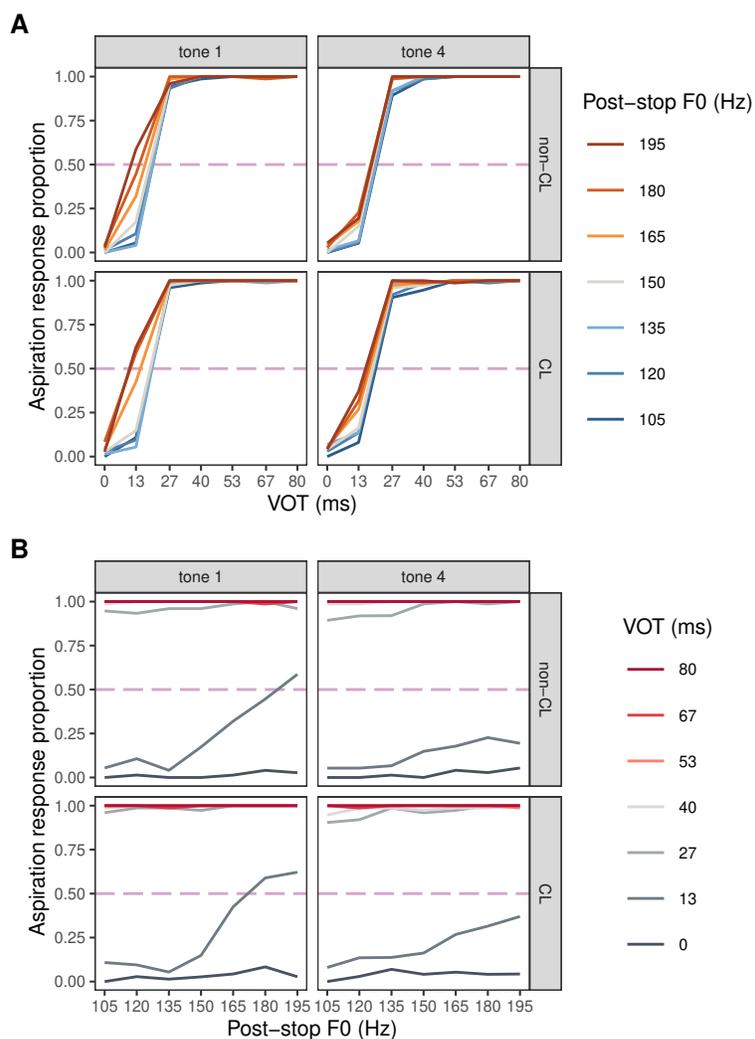
The response patterns over different VOTs, post-stop F0s, tones, and CL conditions are displayed in Figure 5.5 and Figure 5.6 for the within- and between-subject comparisons respectively. The marginal posterior distributions of key population-level parameters are summarized in Table 5.4 and Table 5.5 for the within- and between-subject comparisons respectively. As in the previous chapters, the values of these parameters, with the exception of the category boundary (which is modeled as the intercept in the model), are interpreted as the perceptual weight the listener assigned to each dimension. The distributions of these population-level weights are also visualized in Figure 5.7 to aid interpretations. The weights along these dimensions in both conditions at the individual level are plotted in Figure 5.8 and Figure 5.9. The numerical values behind these figures can be found in Table F.6 (for Figure 5.8), Table F.7 (for Figure 5.9A), and Table F.8 (for Figure 5.9B) in

the appendix. The information about individuals' guessing probabilities is also provided in Figure F.2, Figure F.3, Table F.3 and Table F.4 in the appendix.

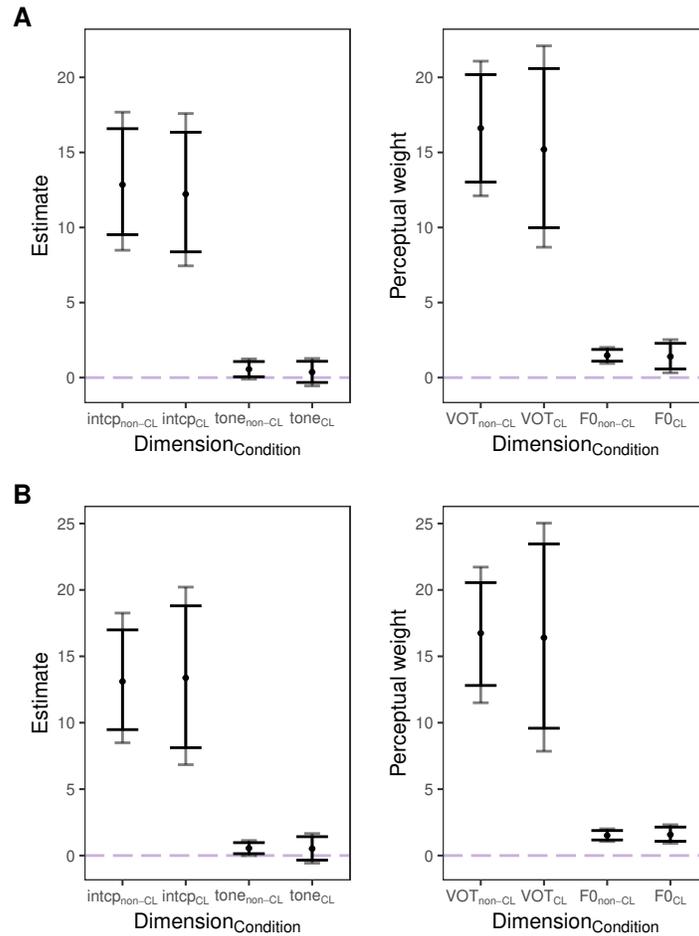


**Figure 5.5:** Aggregated results on English stop-voicing categorization for the within-subject comparison, as a function of VOT, post-stop F0, tone, and CL conditions. **A.** With VOT on the  $x$ -axis. **B.** With post-stop F0 on the  $x$ -axis to highlight its effect on categorization.

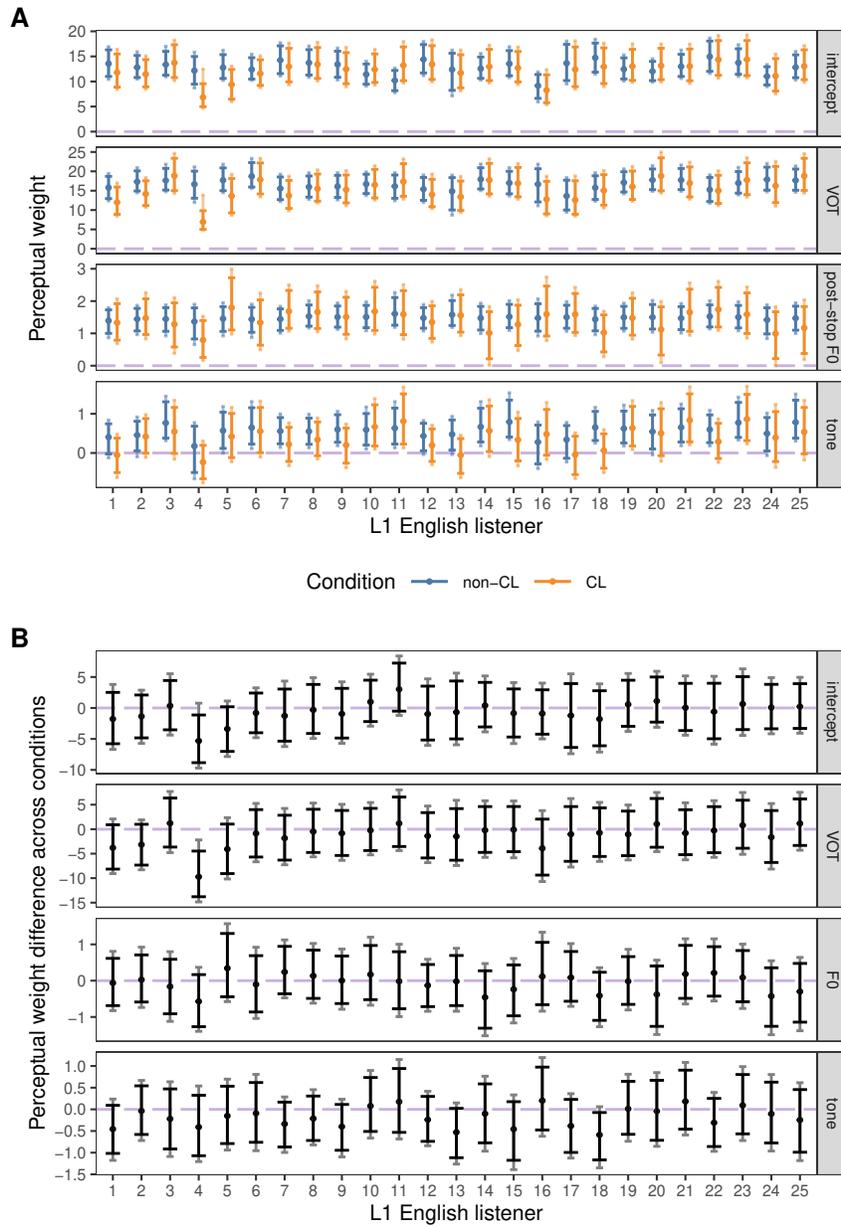
The population-level results reveal that L1 English listeners under the CL con-



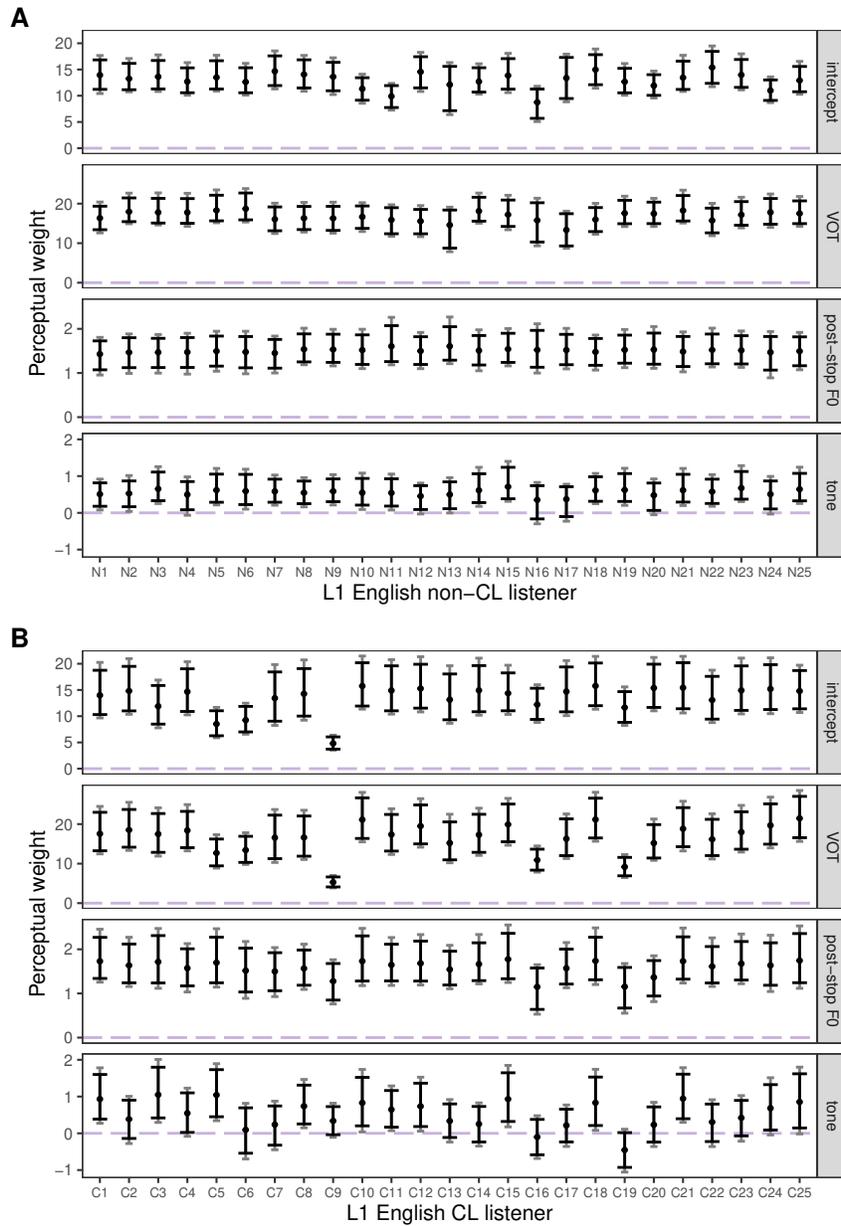
**Figure 5.6:** Aggregated results on English stop-voicing categorization for the between-subject comparison, as a function of VOT, post-stop F0, tone, and CL conditions. **A.** With VOT on the  $x$ -axis. **B.** With post-stop F0 on the  $x$ -axis to highlight its effect on categorization.



**Figure 5.7:** Distributions of perceptual weights along various dimensions at the population level for **A.** the within-subject comparison and **B.** the between-subject comparison. Posterior means are represented by the dots. The 89% CrIs are marked by the inner error bars, and the 95% CrIs are marked by the outer error bars. The CrIs were created from 4,000 samples, each drawn from a normal distribution with the mean and standard deviation corresponding to those from a posterior sample output by the model.



**Figure 5.8:** Individuals' perceptual weights as estimated by the model. **A.** Distributions of individual weights along various dimensions under non-CL and CL conditions. **B.** Differences in cue weights along the same dimension across conditions. Posterior means are represented by the dots. The 89% CrIs are marked by the inner error bars, and the 95% CrIs are marked by the outer error bars.



**Figure 5.9:** Individuals’ perceptual weights estimated by model. **A.** Distributions of individual weights along various dimensions for L1 English listeners under the non-CL condition. **B.** Distributions of individual weights along various dimensions for L1 English listeners under the CL condition. Posterior means are represented by the dots. The 89% CrIs are marked by inner error bars, and the 95% CrIs are marked by outer error bars.

**Table 5.4:** Marginal posterior summary for key parameters from the perceptual model M4 for the English within-subject CL experiment. The symbol  $\mu$  denotes the (distribution of) population mean, and the symbol  $\sigma$  denotes the (distribution of) population standard deviation across participants.

Parameter	Mean	SD	89% CrI	$p(\text{dir.})$
intercept <sub>non-CL</sub> : $\mu$	12.79	.97	[11.33, 14.42]	$p(\beta > 0) = 1.00$
VOT <sub>non-CL</sub> : $\mu$	16.59	1.17	[14.78, 18.51]	$p(\beta > 0) = 1.00$
F0 <sub>non-CL</sub> : $\mu$	1.48	.15	[1.24, 1.72]	$p(\beta > 0) = 1.00$
tone <sub>non-CL</sub> : $\mu$	.56	.12	[.37, .76]	$p(\beta > 0) = 1.00$
intercept <sub>CL</sub> : $\mu$	12.17	1.03	[10.67, 13.94]	$p(\beta > 0) = 1.00$
VOT <sub>CL</sub> : $\mu$	15.25	1.28	[13.37, 17.41]	$p(\beta > 0) = 1.00$
F0 <sub>CL</sub> : $\mu$	1.41	.18	[1.14, 1.70]	$p(\beta > 0) = 1.00$
tone <sub>CL</sub> : $\mu$	.37	.14	[.16, .60]	$p(\beta > 0) = 1.00$
intercept <sub>non-CL</sub> : $\sigma$	1.84	.59	[.92, 2.76]	$p(\beta > 0) = 1.00$
VOT <sub>non-CL</sub> : $\sigma$	1.75	.88	[.25, 3.13]	$p(\beta > 0) = 1.00$
F0 <sub>non-CL</sub> : $\sigma$	.17	.13	[.01, .42]	$p(\beta > 0) = 1.00$
tone <sub>non-CL</sub> : $\sigma$	.26	.15	[.04, .51]	$p(\beta > 0) = 1.00$
intercept <sub>CL</sub> : $\sigma$	2.23	.59	[1.34, 3.20]	$p(\beta > 0) = 1.00$
VOT <sub>CL</sub> : $\sigma$	3.04	.72	[1.99, 4.16]	$p(\beta > 0) = 1.00$
F0 <sub>CL</sub> : $\sigma$	.47	.20	[.13, .79]	$p(\beta > 0) = 1.00$
tone <sub>CL</sub> : $\sigma$	.41	.14	[.20, .64]	$p(\beta > 0) = 1.00$

dition still used VOT as the primary cue and post-stop F0 as a secondary cue. This can be seen in the CL subfigures of Figure 5.5 and Figure 5.6, where the proportion of /p/ response flips from close to 0 to close to 1 as VOT passes over the category boundary, but the effect of post-stop F0 change is most apparent only at the category boundary. However, there is greater variability in the use of all cues in the CL condition. For the VOT cue, the mean weights between the non-CL and CL conditions were similar (within-subject: 89% CrI of  $VOT_{\text{non-CL}} - VOT_{\text{CL}} = [-1.39, 3.93]$ ; between-subject: 89% CrI of  $VOT_{\text{non-CL}} - VOT_{\text{CL}} = [-.64, 1.28]$ ), but the cross-participant standard deviation of the VOT weight in the CL condition tended to be larger than that in the non-CL condition (within-subject: 89% CrI of  $VOT_{\text{non-CL}} - VOT_{\text{CL}} = [-3.11, .49]$ ; between-subject: 89% CrI of  $VOT_{\text{non-CL}} - VOT_{\text{CL}} = [-4.18, -.11]$ ). As for the post-stop F0 cue, CL again did

**Table 5.5:** Marginal posterior summary for key parameters from the perceptual model for the English between-subject CL experiment. The symbol  $\mu$  denotes the (distribution of) population mean, and the symbol  $\sigma$  denotes the (distribution of) population standard deviation across participants.

Parameter	Mean	SD	89% CrI	$p(\text{dir.})$
intercept <sub>non-CL</sub> : $\mu$	13.05	.79	[11.82, 14.34]	$p(\beta > 0) = 1.00$
VOT <sub>non-CL</sub> : $\mu$	16.73	.95	[15.29, 18.31]	$p(\beta > 0) = 1.00$
F0 <sub>non-CL</sub> : $\mu$	1.51	.13	[1.30, 1.72]	$p(\beta > 0) = 1.00$
tone <sub>non-CL</sub> : $\mu$	.55	.10	[.39, .72]	$p(\beta > 0) = 1.00$
intercept <sub>CL</sub> : $\mu$	13.33	.81	[12.10, 14.65]	$p(\beta > 0) = 1.00$
VOT <sub>CL</sub> : $\mu$	16.48	.98	[14.95, 18.05]	$p(\beta > 0) = 1.00$
F0 <sub>CL</sub> : $\mu$	1.58	.13	[1.38, 1.81]	$p(\beta > 0) = 1.00$
tone <sub>CL</sub> : $\mu$	.53	.13	[.33, .73]	$p(\beta > 0) = 1.00$
intercept <sub>non-CL</sub> : $\sigma$	2.09	.66	[1.11, 3.20]	$p(\beta > 0) = 1.00$
VOT <sub>non-CL</sub> : $\sigma$	1.98	1.07	[.24, 3.69]	$p(\beta > 0) = 1.00$
F0 <sub>non-CL</sub> : $\sigma$	.16	.13	[.01, .39]	$p(\beta > 0) = 1.00$
tone <sub>non-CL</sub> : $\sigma$	.20	.14	[.02, .44]	$p(\beta > 0) = 1.00$
intercept <sub>CL</sub> : $\sigma$	3.19	.62	[2.31, 4.23]	$p(\beta > 0) = 1.00$
VOT <sub>CL</sub> : $\sigma$	4.16	.72	[3.08, 5.39]	$p(\beta > 0) = 1.00$
F0 <sub>CL</sub> : $\sigma$	.28	.15	[.05, .54]	$p(\beta > 0) = 1.00$
tone <sub>CL</sub> : $\sigma$	.52	.15	[.29, .76]	$p(\beta > 0) = 1.00$

not seem to have an effect on the means (within-subject: 89% CrI of  $F0_{\text{non-CL}} - F0_{\text{CL}} = [-.30, .43]$ ; between-subject: 89% CrI of  $F0_{\text{non-CL}} - F0_{\text{CL}} = [-.34, .14]$ ) but did appear to slightly increase the standard deviation (within-subject: 89% CrI of  $F0_{\text{non-CL}} - F0_{\text{CL}} = [-.68, .09]$ ; between-subject: 89% CrI of  $F0_{\text{non-CL}} - F0_{\text{CL}} = [-.44, .17]$ ). These results are therefore at odds with Gordon et al. (1993), which predicts the VOT weight to decrease and the post-stop F0 weight to increase in the CL condition. On the other hand, it seems that these results agree with what has been observed in Mattys and Wiget (2011)—a lack of change in the VOT weight—though they did not look into the change in variation of the weight.

The enlarged population-level variation in both VOT and post-stop F0 weights under the CL condition translates to more between-individual variability at the individual level. This increased between-individual variability is visible in both

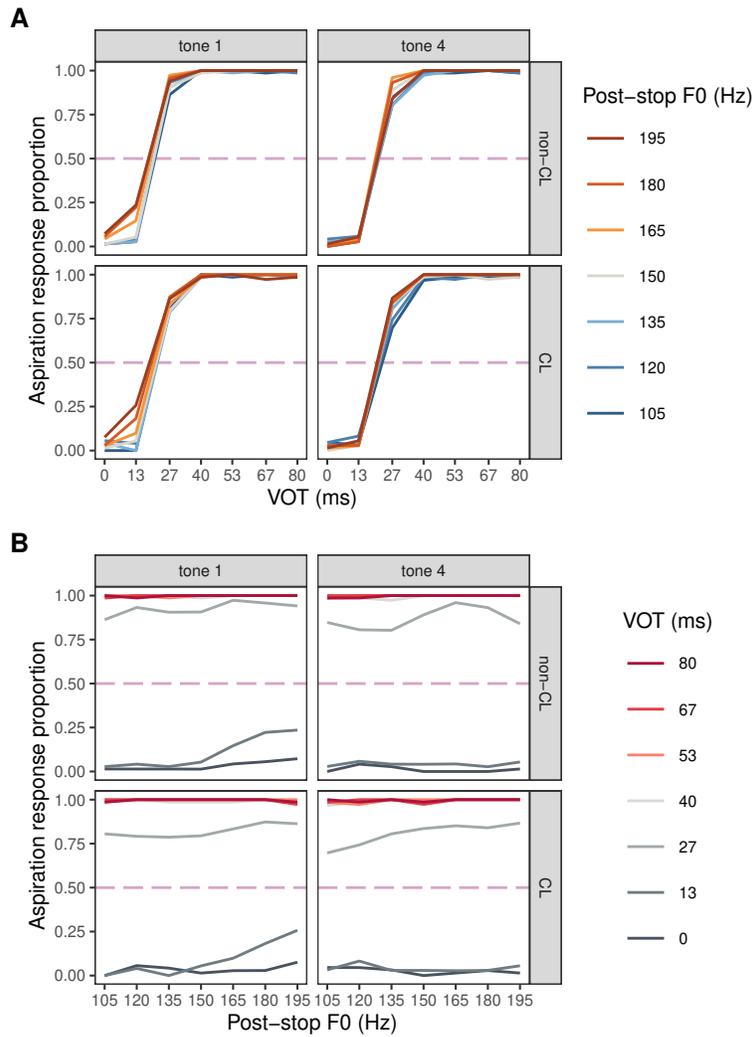
Figure 5.8A and Figure 5.9, where the posterior means show more fluctuation in the CL condition. *Within*-individual variability in both weights also increased under the CL condition. This is easiest to spot in Figure 5.8A, where the 89% CrIs span a larger range under the CL condition for both the individual VOT and, in particular, individual post-stop F0 weights. However, similar to the results at the population level, individual listeners did not seem to systematically raise or lower perceptual weights, either for VOT or post-stop F0, under the CL condition. This can be seen clearly in Figure 5.8B, where the 89% CrIs for the difference in cue weights along all dimensions cross zero for almost all listeners.

In short, the English CL experiment show that, contra what has been found in previous studies, the CL manipulation did not change the overall perceptual weights of the primary (i.e., VOT) and a secondary (i.e., post-stop F0) cue at either the population or individual level. However, CL did seem to bring about more between- and within-individual variabilities in cue weights.

### **Mandarin Between-Subject non-CL versus CL Comparison**

The response patterns from L1 Mandarin listeners under non-CL and CL conditions are plotted in Figure 5.10. The marginal posterior distributions of cue weights at the population level are tabulated in Table 5.6 and visualized in Figure 5.11. Individual-level cue weights are visually summarized in Figure 5.12, and the precise values are listed in Table F.9 and Table F.10 in the appendix. Individual listeners' estimated guessing probabilities are also provided in Figure F.4 and Table F.5 in the appendix.

The population-level results demonstrate that L1 Mandarin listeners under the CL condition still used both VOT and post-stop F0 as cues when categorizing voicing of stops in their native language. This finding is similar to the English results, but there is no similar across-the-board increase in variability in the cues. Specifically, for the VOT cue, the mean weights stayed largely the same across non-CL and CL conditions (89% CrI of  $VOT_{\text{non-CL}} - VOT_{\text{CL}} = [-1.43, .61]$ ), and the same could be stated for the mean post-stop F0 weights (89% CrI of  $F0_{\text{non-CL}} - F0_{\text{CL}} = [-.54, .05]$ ). CL, however, seemed to have a *shrinking* effect on the standard deviation of the VOT cue (89% CrI of  $VOT_{\text{non-CL}} - VOT_{\text{CL}} = [-.92, 3.92]$ ), but not on



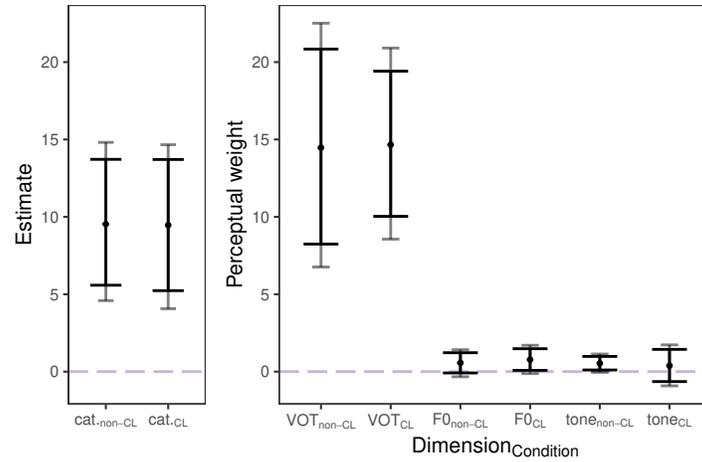
**Figure 5.10:** Aggregated results on Mandarin stop-voicing categorization, as a function of VOT, post-stop F0, tone, and CL conditions. **A.** With VOT on the  $x$ -axis. **B.** With post-stop F0 on the  $x$ -axis to highlight its effect on categorization.

**Table 5.6:** Marginal posterior summary for key parameters from the perceptual model for the Mandarin between-subject CL experiment. The symbol  $\mu$  denotes the (distribution of) population mean, and the symbol  $\sigma$  denotes the (distribution of) population standard deviation across participants.

Parameter	Mean	SD	89% CrI	$p(\text{dir.})$
intercept <sub>non-CL</sub> : $\mu$	9.47	.59	[8.58, 10.45]	$p(\beta > 0) = 1.00$
VOT <sub>non-CL</sub> : $\mu$	14.43	.95	[12.98, 15.99]	$p(\beta > 0) = 1.00$
F0 <sub>non-CL</sub> : $\mu$	.57	.14	[.34, .79]	$p(\beta > 0) = 1.00$
tone <sub>non-CL</sub> : $\mu$	.54	.12	[.36, .73]	$p(\beta > 0) = 1.00$
intercept <sub>CL</sub> : $\mu$	9.42	.63	[8.44, 10.46]	$p(\beta > 0) = 1.00$
VOT <sub>CL</sub> : $\mu$	14.67	.91	[13.30, 16.21]	$p(\beta > 0) = 1.00$
F0 <sub>CL</sub> : $\mu$	.78	.15	[.55, 1.02]	$p(\beta > 0) = 1.00$
tone <sub>CL</sub> : $\mu$	.40	.16	[.14, .64]	$p(\beta > 0) = .99$
intercept <sub>non-CL</sub> : $\sigma$	2.38	.53	[1.60, 3.29]	$p(\beta > 0) = 1.00$
VOT <sub>non-CL</sub> : $\sigma$	3.81	.80	[2.66, 5.22]	$p(\beta > 0) = 1.00$
F0 <sub>non-CL</sub> : $\sigma$	.36	.15	[.14, .61]	$p(\beta > 0) = 1.00$
tone <sub>non-CL</sub> : $\sigma$	.22	.13	[.03, .45]	$p(\beta > 0) = 1.00$
intercept <sub>CL</sub> : $\sigma$	2.51	.57	[1.70, 3.50]	$p(\beta > 0) = 1.00$
VOT <sub>CL</sub> : $\sigma$	2.38	1.28	[.22, 4.26]	$p(\beta > 0) = 1.00$
F0 <sub>CL</sub> : $\sigma$	.40	.16	[.17, .65]	$p(\beta > 0) = 1.00$
tone <sub>CL</sub> : $\sigma$	.63	.16	[.39, .91]	$p(\beta > 0) = 1.00$

that of the post-stop F0 cue (89% CrI of  $F0_{\text{non-CL}} - F0_{\text{CL}} = [-.40, .28]$ ), much to the contrary of its uniformly deviation-*expanding* effect just seen in English.

At the individual level, the effect of CL also appeared to be limited. Specifically, except for reducing between-individual variation in the VOT cue (this is just a restatement of the standard deviation being shrunk for the VOT cue at the population level), CL seemingly did not trigger other changes in VOT or post-stop F0 weights.

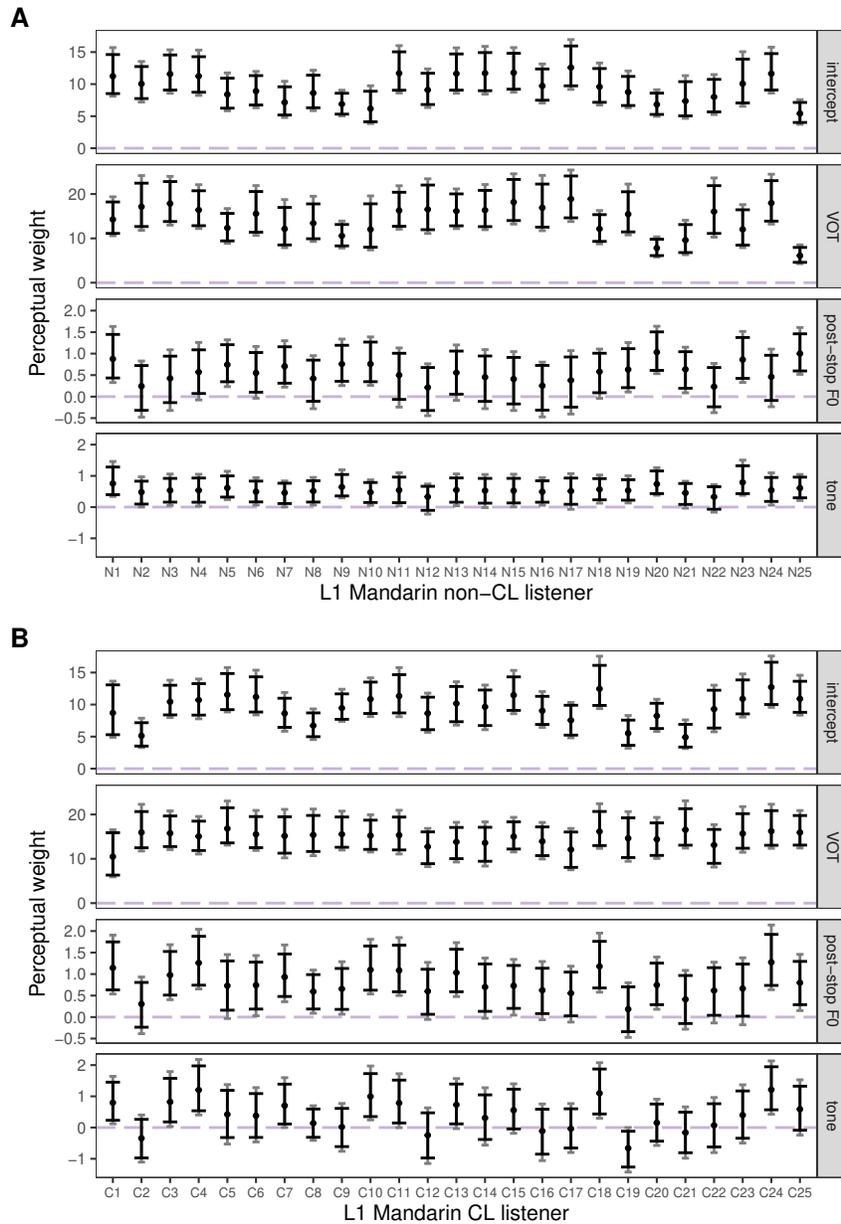


**Figure 5.11:** Distributions of perceptual weights along various dimensions at the population level. Posterior means are represented by the dots. The 89% CrIs are marked by the inner error bars, and the 95% CrIs are marked by the outer error bars. The CrIs were created from 4000 samples, each drawn from a normal distribution with the mean and standard deviation corresponding to those from a posterior sample output by the model.

## 5.3 Discussion

### 5.3.1 Summary of Results

This chapter examines the effect of CL on L1 English and L1 Mandarin listeners' use of VOT and post-stop F0 in perceiving the voicing of stops in their respective L1s. The experimental results failed to replicate the patterns described in Gordon et al. (1993). That is, instead of lowering the weight of VOT and raising the weight of post-stop F0, the manipulation of CL did not seem to shift the average weight of either dimension. This is not to say that CL did not have any impacts. Indeed, CL appeared to modulate the variation or uncertainty in the estimated weight values (cf. Ciaccio and Veríssimo [2020] for a similar finding in the domain of morphological processing). However, even in this aspect, the effect seemed to be language-dependent. In particular, while CL increased the uncertainty around both the VOT and post-stop F0 weights for L1 English listeners in both the within- and



**Figure 5.12:** Individuals’ perceptual weights estimated by model. **A.** Distributions of individual weights along various dimensions for L1 Mandarin listeners under the non-CL condition. **B.** Distributions of individual weights along various dimensions for L1 Mandarin listeners under the CL condition. Posterior means are represented by the dots. The 89% CrIs are marked by inner error bars, and the 95% CrIs are marked by outer error bars. 182

between-subject comparisons, it seemed to only shrink the cross-participant standard deviation of the VOT cue for L1 Mandarin listeners, for which there was only a between-subject comparison.

The experimental results are juxtaposed with the expected results in Table 5.7. These comparisons beg the question of why CL seems to impose mildly diverging effects on the variability of cues across studies. The observation in the current study that CL is linked to increased variability also demands some explanations. In what follows I speculate on why the current work did not replicate the findings from Gordon et al. (1993) and on the causes of the differential effects CL has on L1 English and L1 Mandarin listeners.

**Table 5.7:** Predicted and actual results under different hypotheses.

Hypothesis	Predicted results	Match experiment results?
The visual search task and the arithmetic task tap into the same central processing resources, and the post-stop F0 manipulation in this study has a similar effect as the overall F0 manipulation in Gordon et al. (1993).	VOT weight in CL ↓, and post-stop F0 weight in CL ↑ for all three comparisons.	✗ Mean VOT and post-stop F0 weights did not change under CL for either language. Instead, CL added uncertainty to estimated weights, at least for English listeners.
The visual search task has no effect on cue weighting, as hinted on in Mattys and Wiget (2011), or the post-stop F0 manipulation in this study has a different effect than the F0 manipulation in Gordon et al. (1993). Alternatively, it could simply be that the findings in Gordon et al. (1993) are not replicable.	Weight in CL $\approx$ weight in non-CL for VOT and post-stop F0 in all three comparisons.	✓ <b>VOT:</b> Mean VOT and post-stop F0 weights did not change under CL for either language. <b>Variability:</b> CL added uncertainty to individual VOT and post-stop F0 weights, but only for the L1 English group.

### 5.3.2 Cognitive Load Experiments: Same, Same but Different

As pointed out in Section 5.1.1, the conducted experiments differ from the one documented in Gordon et al. (1993) in two important aspects: the nature of the CL task and the way F0 was manipulated. Given that different results were obtained, it is important to consider how these two aspects might affect the outcomes.

First, the differences might have arisen from some CL tasks inducing verbal rehearsal during the task. Recall that a visual search task was adopted as opposed to an arithmetic task originally used in Gordon et al. (1993). In the arithmetic task, listeners were presented a series of three numbers and had to perform a simple calculation with the numbers. The presence of the numbers and the required operation might have prompted the listeners to mentally rehearse the numbers linguistically. This might in turn cause the secondary task (i.e., numerical operations in the arithmetic task) to interfere with the primary task (i.e., phoneme classification) in ways not intended (Park and Brünken, 2015). In short, the apparent CL effect might not be purely due to the depletion of central processing resources per se, but an implicitly-induced verbal rehearsal could be at work too. A related issue is that the two types of tasks might not distract the listener to a similar extent. For instance, the arithmetic task might compete for more attention than the visual search task (i.e., the mental arithmetic task requires the participant to hold numbers in the working memory when performing mathematical operations, while the visual search task has no such requirement), and the kind of cue-weight shift observed in Gordon et al. (1993) might be contingent on the decrease in selective attention being substantial.

Second, different ways of F0 manipulation might affect how the F0 cue was used by listeners. The F0 manipulation in Gordon et al. (1993) involved only two levels (i.e., 100 Hz and 180 Hz) and spanned the entire vowel portion, whereas the F0 manipulation in this chapter consisted of seven incremental steps confined to the initial 35% of the vowel. In addition, the overall tonal contour of a stimulus was manipulated. While a robust effect of F0 was observed under no load in both cases, it might be the case that the additional variation in F0 (introduced by both the vowel-initial F0 and overall tonal contour manipulations) in this study masks certain effects of CL. Of course whether these factors are at play await further research.

### **5.3.3 Language-Specific Effects of Cognitive Load?**

Another issue that needs to be addressed concerns the effect of CL on the variability/uncertainty of estimated weights and its seemingly very mildly language-

specific impacts (i.e., the outcome was largely the same across the non-CL and CL conditions, with the exception of small differences in the cross-participant standard deviation of the VOT and post-stop F0 cues). For explanations, I resort to one of the proposed mechanisms that underlies CL's influence on phonetic encoding. The mechanism in question suggests that CL adversely affects speech perception due to a decrease in the perceptual *signal-to-noise* ratio (Gordon et al., 1993; Mattys and Wiget, 2011). In other words, CL induces a drop in the strength of faithful encoding of incoming speech signal compared to the level of the background system noise. This suboptimal signal-to-noise ratio could be the consequence of a reduction in the processing strength of speech cues and/or a deterioration in filtering out system noise. As this mechanism assumes random noise, CL would be expected to impose a general negative impact on the encoding of all phonetic cues, therefore including VOT and post-stop F0.

This less accurate encoding of VOT and post-stop F0 under CL means that both cues become less reliable in indicating the phonemic membership of the perceived speech signal. In terms of the response patterns in the data, there would be more random fluctuations in categorization. In turn, these random fluctuations translate to increased uncertainty around estimated weights—this is what has been observed for both groups of L1 English listeners.

However, this increase-in-uncertainty trend is absent in L1 Mandarin listeners' data. Three speculative factors might be responsible. First, looking at the panel corresponding to VOT in Figure 5.12A, it seems that the large population-level variation in VOT in the non-CL condition (see Figure 5.11) was mainly driven by three listeners (i.e., N20, N21, and N25) who had exceptionally low VOT weights in comparison with the other listeners. It is therefore possible that the apparent 'shrinking' effect of CL on L1 Mandarin listeners was an artifact of the fact that the listeners in the non-CL condition happened to have greater between-individual variation. Running the same experiment on another group of listeners will be able to clarify this point. Second, the lack of effect on the post-stop F0 cue might be attributed to the fact that post-stop F0 is inherently a weak cue for stop voicing in L1 Mandarin listeners, as compared to L1 English listeners. It is possible that CL has little or limited impact on a weak perceptual cue. To verify this claim, we first need to ensure that this absence of effect is consistent across multiple iterations of

the same experiment and then follow up with a systematic cross-linguistic investigation. Finally, the lack of CL effect in Mandarin listeners could simply be due to their not paying as much attention in the distractor task as the English groups. As a result, it is not surprising that the manipulation of CL contributes little to their performance. This can be justified partially by comparing Mandarin listeners' accuracy in the visual search task (Figure 5.4C) with that of English listeners (Figure 5.4A and Figure 5.4B): while both groups showed the same improvement pattern, Mandarin listeners as a whole were lagging behind English listeners across the three blocks (i.e., the mean accuracy of each block for the two groups of L1 English listeners was (.75, .77, .80) and (.75, .79, .81), while that for the L1 Mandarin listeners was (.69, .73, .76)). One way to address this problem is to run more participants and only include participants with similar CL accuracy in the analyses.

## 5.4 Conclusion

The present work examined whether and how cognitive load interacted with L1 English and L1 Mandarin listeners' use of the VOT and post-stop F0 cues in stop-voicing classification. The additional cognitive load was introduced by having the listener perform a concurrent visual search task while hearing the audio stimuli. Statistical analyses suggest that cognitive load affected the two listener groups asymmetrically. For English listeners, contra previous findings, cognitive load did not seem to systematically raise or lower either the VOT or post-stop F0 weight. However, cognitive load put more uncertainty around each estimated weight. The failure to replicate previous findings might be due to two critical design choices adopted in this study. Cognitive load's uncertainty-expanding property is potentially linked to its detrimental effect on phonetic encoding. For Mandarin listeners, cognitive load appeared to have little impact. This absence of effect is speculatively ascribed to factors related to chance participant configuration, post-stop F0 use in Mandarin, and inattentiveness during the distractor task.

## Chapter 6

# General Discussion

### 6.1 Summary of Main Findings

The corpus study and the set of parallel experiments presented in this work provide insight into how speakers and listeners of Mandarin and English make use of the post-stop cue when producing and perceiving native and foreign stop manner contrasts. One set of experiments was also designed to address the question of whether and how listeners modify their cue weights when confronted with additional cognitive load.

Chapter 2 presented a corpus study on the impact of phonological voicing— aspirated, unaspirated, and sonorant—of the initial consonant on vowel-onset fundamental frequency (F0) in Mandarin production. Two methods of F0 normalization—F0 standardization and F0 scaling—were explored, with each method capturing a different profile of an F0 trajectory at vowel onset. Specifically, F0 standardization, where vowel-onset F0 is standardized/*z*-transformed within individual speakers, captures and maintains the relative ordering of vowel-onset F0 in the raw data. On the other hand, F0 scaling, where vowel-onset F0 is scaled relative to the mean F0 of initial portion of the F0 trajectory, characterizes how steeply an F0 trajectory drops within the initial portion of a vowel. Altogether, the results of this corpus study paint a picture where aspirated stops in Mandarin tend to have a lower vowel-onset F0 than unaspirated stops (based on the standardization method), though the effect size, as evaluated by Cohen's *d* ( $\bar{d} = .03$ ), is fairly

small. On the other hand, the F0 trajectory following an aspirated stop generally has a sharper drop than that following an unaspirated stop (based on the scaling method).

Chapter 3 provided a comparison of production and perception of post-stop F0 in stop voicing contrasts in two languages, Mandarin and English, by L1 (first language) Mandarin-L2 (second language) English speakers. Given that F0 is the primary acoustic correlate of lexical tones in Mandarin, Mandarin provides an opportunity to examine whether F0 still functions as a marker for stop voicing in production and as a cue for stop voicing in perception when the bilingual listeners also need to extract tonal information from F0. A related question is whether the use of post-stop F0, in both production and perception, is sensitive to different language contexts. In the bilinguals' production of both Mandarin and English, using *z*-score normalization, the post-stop F0 following an aspirated stop tended to be higher than that following an unaspirated stop. In addition, the degree to which post-stop F0 was differentiated between the aspirated and unaspirated series depended on the language, such that English displayed a larger post-stop F0 difference than Mandarin. In perception, post-stop F0 was also used as a cue for stop voicing when the same bilinguals were put in either a Mandarin or an English context. Furthermore, the language context modulated the perceptual weight: post-stop F0 carried more weight when the stimuli were presented as English words than when the acoustically identical stimuli were presented as Mandarin words. While the same general patterns held for production and perception in terms of the use of post-stop F0, attempts at linking production and perception at the individual level were only partially successful: there was a positive correlation between production and perceptual post-stop F0 weights in English, but not in Mandarin. However, even this positive correlation should be interpreted with caution as uncertainty around cue weights was not taken into account.

Chapter 4 used similar tasks to compare L1 Mandarin-L2 English bilinguals' use of post-stop F0 in English production and perception with that of L1 English speakers. In production, the bilingual speakers' use of post-stop F0 cue patterned with L1 English speakers in terms of both direction and magnitudes: the F0 following a voiceless/aspirated stop tended to be higher than that following a voiced/unaspirated one, and the two speaker groups' post-stop F0 weights in production were

similar to each other. Both groups also used post-stop F0 as a cue for stop voicing in perception; however, there was weak evidence that, on average, bilingual listeners relied on post-stop F0 cue less than L1 English listeners. In addition, when the same individual's mean production and perceptual weights were correlated, there was a weak positive correlation for both groups. As above, though, this apparent positive relationship across modalities might be overconfident as uncertainty was ignored in the model.

Finally, chapter 5 switched gears a bit and explored the effect of cognitive load on L1 English and L1 Mandarin listeners' use of voice onset time (VOT) and post-stop F0 cues in perceiving the voicing of stops in their respective L1. Based on the findings from Gordon et al. (1993), the weight for VOT, being the primary cue, was expected to decrease while the weight for post-stop F0, being a secondary cue, was predicted to increase. These patterns were not observed—instead the manipulation of cognitive load did not seem to shift the average perceptual weight of either dimension. However, cognitive load appeared to modulate variance in the estimated weight values, though this variance-modulating effect was language-dependent. Specifically, while cognitive load led to increased variance for both VOT and post-stop F0 weights for L1 English listeners, it did not affect the weights of the two cues much for L1 Mandarin listeners.

## 6.2 Revisiting the Research Questions

Having summarized the main findings from individual chapters, this section ties these findings back to the research questions this work set out to answer:

- **RQ1:** Is post-stop F0 in L1 Mandarin speakers' production correlated with the phonological voicing of stops in the language?
  - The corpus study in Chapter 2 suggests a positive answer: the post-stop F0 at vowel onset after an aspirated stop is on average lower than that after an unaspirated stop.
  - The results from the Mandarin production experiment in Chapter 3 are also affirmative. However, the statistical model supports the opposite pattern as the corpus study, with the post-stop F0 after an aspirated stop

being on average *higher* than that after an unaspirated stop.

- **RQ2:** How do L1 Mandarin-L2 English bilinguals weight post-stop F0 in their *production* of Mandarin and English stops? Does the language context play a role in their use of post-stop F0 in production?
  - The production data from the paired experiments in Chapter 3 shows that, in both their Mandarin and English tokens, the bilinguals produce aspirated stops with a higher post-stop F0 than unaspirated stops. In addition, the language context has an effect such that English tokens display a larger post-stop F0 difference than Mandarin tokens.
- **RQ3:** Relatedly, how do L1 Mandarin-L2 English bilinguals weight post-stop F0 in their *perception* of Mandarin and English stops? Again, does the language context modulate the use of post-stop F0 in perception?
  - The identification data from the perception experiments in Chapter 3 suggests that post-stop F0 is used as a cue for stop voicing in both Mandarin and English contexts. However, the language context modulates the reliance on post-stop F0 for voicing: that is, post-stop F0 carries more weight when the stimuli are presented as English words than when they are presented as Mandarin words.
- **RQ4:** How does L1 Mandarin-L2 English bilinguals' production and perceptual use of post-stop F0 as a cue for the stop voicing in English compare to that of L1 English speakers?
  - The English production data from Chapter 4 indicates that the bilinguals pattern with L1 English speakers such that the F0 following an aspirated stop is likely to be higher than that following an unaspirated stop, and the post-stop F0 difference between the two stop class is similar across the two groups. The English perceptual data suggests that both groups also use post-stop F0 as a cue for stop voicing, but L1 English listeners assign more weight to this cue than bilingual listeners. This asymmetry between production and perception is taken up again in Section 6.4.

- **RQ5:** On both the population and individual levels, is the production weight of post-stop F0 predictive of the perceptual weight of the same cue?
  - On the group level, it is generally observed that a higher production weight of post-stop F0 is indicative of a higher perceptual weight of post-stop F0, and vice versa. On the individual level, there seems to be a weak correlation between production weight and perceptual weight in L1 and L2 English. However, given that uncertainty around weight estimates is not taken into account, the actual correlation is likely to be even weaker.
- **RQ6:** How does cognitive load affect the use of post-stop F0 in identifying stop voicing in Mandarin and English?
  - The within- and between-subject comparisons across load and non-load identification data suggest that cognitive load does not shift the mean perceptual weight of post-stop F0, but it increases its variance for L1 English listeners.

### **6.3 Reconciling the Differences between the Corpus and Experimental Studies**

As just summarized above, the corpus study in Chapter 2 shows a post-stop F0 pattern that conflicts with what has been observed in the controlled production experiment in Chapter 3. In particular, while the corpus study suggests that post-stop F0 after an aspirated stop is lower than that after an unaspirated stop, the data from the production experiment reveals the opposite pattern. In what follows, I offer a speculative explanation on the cause that might give rise to these diverging results.

One possible explanation is cross-linguistic influence (CLI). CLI is an integrated component of the Speech Learning Model (SLM; Flege, 1995, 2007) and the Revised Speech Learning Model (SLM-r; Flege and Bohn, 2021). As the main theoretical tenets of the SLM have been described in Section 4.1.1, the reader is referred to that section for a general overview of the model. Crucial for the current

discussion is that CLI is bidirectional: CLI is not limited to L1 phones affecting L2 phones (“forward transfer”), but may also come from L2 phones influencing L1 phones (“backward transfer”). Such bidirectional influences have been documented in a number of studies, with the investigated units ranging from VOT (e.g., Bullock and Toribio, 2009; Fricke et al., 2016), to vowels (e.g., Guion, 2003), to laterals (e.g., Amengual, 2018; Barlow, 2014), and to fricatives (e.g., Peng, 1993). For instance, in a corpus study, Fricke et al. (2016) demonstrate that Spanish-English bilinguals shorten their VOTs, which makes them more Spanish-like, in the lead up to an English-to-Spanish code-switching point. Flege (1987) reports a case where many L1 English learners of French produce French /t/ with “too-long” VOTs (i.e., as influenced by English) and English /t/ with “too-short” VOTs (i.e., as influenced by French); a similar pattern is also observed for the VOTs from L1 French learners of English. In the current work, the more English-like post-stop F0 pattern (i.e., F0 higher after aspirated stops than unaspirated ones) in L1 Mandarin-L2 English bilinguals’ Mandarin production might be due to L2-to-L1 transfer of the use of post-stop F0. If this transfer was not present for the Mandarin speakers in the corpus study, this could explain their different post-stop F0 pattern. Of course, given that language background of the speakers contributing to the corpus is not available, it is not possible to establish whether those speakers are monolingual Mandarin speakers, and by extension whether their post-stop F0 pattern is representative of monolingual speakers. To test this L2-to-L1 backward transfer CLI explanation therefore requires collecting speech data from monolingual Mandarin speakers.

In closing this section, it is worth pointing out that Mandarin is not the only tonal language for which conflicting post-stop F0 profiles have been observed. Similar diverging results have also been reported for Cantonese, where Francis et al. (2006) show an unaspirated > aspirated pattern, but Ren and Mok (2021) report an aspirated > unaspirated pattern. Ren and Mok (2021) likewise attribute their pattern to English transfer effects.

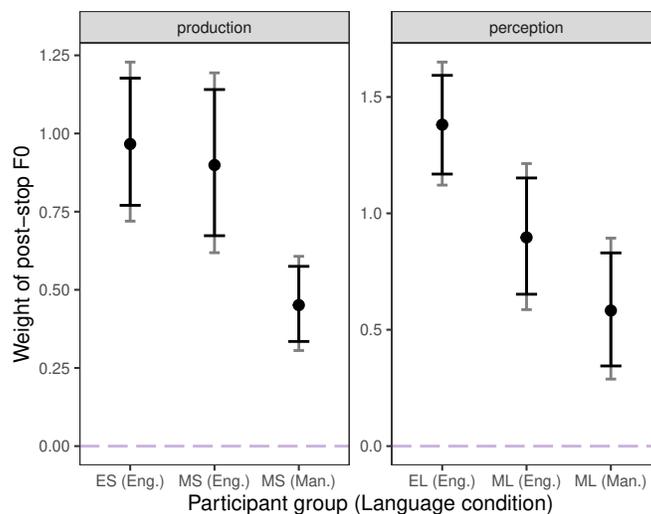
## 6.4 Asymmetry between Production and Perception

The sets of parallel experiments in Chapter 3 and Chapter 4 provide an opportunity to examine a number of issues related to the production-perception interface. Among the broader implications discussed in each chapter independently, the findings of the combined work are particularly relevant to discussions of the interaction between the production-perception interface and L1 influence, as well as between the production-perception interface and individual differences.

### 6.4.1 The Production-Perception Interface and L1 Influence

The production and perception data from Chapter 3 and Chapter 4 allow for a direct comparison of production and perceptual weights of post-stop F0 in three conditions—L1 English speakers in English, L1 Mandarin-L2 English bilinguals in English, and L1 Mandarin-L2 English bilinguals in Mandarin. The population-level production weights under the three conditions are plotted in the left panel of Figure 6.1. These weights were directly taken from Chapter 3 and Chapter 4. The population-level perceptual weights under the three conditions are juxtaposed in the right panel of Figure 6.1. Instead of taking the perceptual weights directly from Chapter 3 and Chapter 4, the perceptual data from the bilinguals in the Mandarin and English contexts as well as that from the L1 English listeners in the non-CL condition were combined and analyzed together to derive the perceptual weight under each condition. Using a single model to estimate weights permits the uncertainty around each weight to be properly handled. The model structure was similar to that used in Chapter 3 and Chapter 4, and the formal specification and output of the model can be found in Section G.1 and Table G.1 in the appendix.

As can be seen in Figure 6.1, there are both symmetric and asymmetric patterns across modalities. In particular, when focusing on L1 Mandarin-L2 English bilinguals' production and perception of the voicing contrast in Mandarin and English, a symmetric pattern can be discerned. That is, a higher production weight of post-stop F0 in English, as compared to Mandarin, is mapped to a higher perceptual weight of the same cue in English. However, when focusing on the English performances from L1 English and L1 Mandarin-L2 English bilinguals, an asymmetric pattern is observed. Specifically, while the two groups share similar production



**Figure 6.1:** Comparison of production and perceptual weights of post-stop F0 under different conditions. Note that the production and perceptual weights are on difference scales, so they are not directly comparable with each other. ES = L1 English speaker, MS = L1 Mandarin-L2 English speaker, EL = L1 English listeners, ML = L1 Mandarin-L2 English listener, Eng. = English task, and Man. = Mandarin task.

weights, L1 English listeners rely on post-stop F0 more than bilingual listeners in perception.

As discussed in Section 3.5.2, the bilinguals' symmetric pattern across modalities is a testament to bilinguals' ability in adapting cue-weighting strategies according to the language context. However, this flexibility in shifting cue weights is not without constraints, as signified by the asymmetric pattern across modalities when L1 English speakers' and L1 Mandarin speakers' English performances are compared. In particular, even though L1 Mandarin listeners already assign a higher perceptual weight to post-stop F0 in the English context, L1 English listeners have the highest weight for post-stop F0. L1 Mandarin listeners' lower perceptual weight for post-stop F0 is likely due to the influence from Mandarin, where the processing dependency between segmental and tonal information, as well as L1 listeners being sensitive to irrelevant tonal variation, lead to post-stop F0 being a less reliable cue for voicing in L1 Mandarin listeners than in L1 English listeners (see

Section 4.1.3). Post-stop F0 being a weak cue for voicing in L1 Mandarin listeners might also be tied to the fact that previous studies have reported mixed results on the directionality of post-stop F0 perturbation in Mandarin. Indeed, conflicting results have been observed in the current work: the corpus study in Chapter 2 has 19 out of 20 L1 speakers who had a *lower* mean post-stop F0 for aspirated stops than for unaspirated ones, but the production experiment in Chapter 3 has all 25 L1 speakers having a *higher* mean post-stop F0 for aspirated stops than for unaspirated ones. L1 Mandarin listeners might therefore assign much less importance to this dimension because of its lack of informativeness about voicing.

The attribution of differential post-stop F0 weights in perception to L1 influence still leaves open the question of why there seems to be less L1 carryover in L1 Mandarin speakers' English production, at least in terms of the chosen measure of post-stop F0 weight in production. In other words, if L1 influence is across the board, we would expect the production weight for post-stop F0 by L1 Mandarin speakers in the English context to be between that by L1 English speakers and that by L1 Mandarin speakers in the Mandarin context. However, L1 Mandarin speakers' population-level production weight is on par with L1 English speakers' production weight. In this connection, two questions regarding the mechanism(s) behind post-stop F0 perturbation are relevant. First, is post-stop F0 perturbation induced by an automatic process attributable to physiological/aerodynamic reason (e.g., Halle and Stevens, 1971; Hombert et al., 1979; Kohler, 1984; Ladefoged, 1967; Ohala and Ohala, 1972), a controlled process—either conscious or subconscious (Kingston and Diehl, 1994), or a combination of the two? Second, if a controlled process is part of the mechanism, do the speakers of tonal languages actively inhibit post-stop F0 perturbation or do the speakers of non-tonal languages actively exaggerate post-stop F0 perturbation? The findings here seem to be more consistent with a combined mechanism and the inhibition hypothesis. To elaborate, if post-stop F0 perturbation has its origin in an automatic process, it is anticipated that the English production from L1 English and L1 Mandarin speakers would have similar post-stop F0 differences—and therefore similar production weights—which is what has been observed. However, the differential post-stop F0 difference across English and Mandarin by the same L1 Mandarin speakers is aligned with the idea that post-stop F0 perturbation is also a controlled process, such that the

pressure to maintain distinctive tonal contours may prompt L1 Mandarin speakers to reduce the extent of post-stop F0 perturbation in Mandarin. In short, L1 Mandarin speakers have an automatic post-stop F0 perturbation effect in their L2 English but a suppressed post-stop F0 perturbation process in their L1 Mandarin. Of course the argument above does not directly contradict the exaggeration hypothesis: L1 English speakers (and even L1 Mandarin speakers when producing English words) may exaggerate post-stop F0 difference to enhance the percept of voicing (Keyser and Stevens, 2006; Kingston, 2007; Kingston and Diehl, 1994). Further research that simulates the perturbation effect with aerodynamic and physiological models can help illuminate this debate. For instance, by comparing the size of the simulated perturbation effect with that of the actual effect observed in different languages, if the size or temporal extent of the actual effect in non-tonal languages is larger than the simulated effect, it is supportive of the exaggeration hypothesis.

In sum, the symmetric as well as asymmetric patterns between production and perception revealed in the present cross-linguistic investigation attests both bilinguals' flexibility in shifting cue weights and influences of L1 on L2 perception. In addition, comparing production from L1 English and L1 Mandarin speakers suggests that post-stop F0 perturbation is rooted in a combined mechanism, where the perturbation effect is attenuated in L1 Mandarin speakers' Mandarin production in order to sustain tonal contrast.

#### **6.4.2 Individual Differences in Production and Perception**

A common thread running through each experiment in this work is a focus on individual variability, which was explored both in L1 English and L1 Mandarin production and perception (in Chapter 2, Chapter 3, and Chapter 4) and in the production and perception of L2 English contrast by L1 Mandarin speakers (in Chapter 4). The fact that individuals vary in the use of the same cue raises important questions about the motivations behind such variation. In this section, I discuss the connection, or lack thereof, between individual differences in cue use and three mechanisms that have been proposed to account for these differences—speaker experience (e.g., Baker and Trofimovich, 2005; Bohn and Flege, 1997; Flege et al.,

1997, 1996; Gorba and Cebrian, 2021; Lev-Ari, 2018; Lev-Ari and Peperkamp, 2016), language dominance (e.g., Casillas, 2015; Casillas and Simonet, 2016; Dmitrieva, 2019; Hazan and Boulakia, 1993), and production-perception interface (e.g., Fowler, 1986; Liberman and Mattingly, 1985).

### **Speaker Experience**

A primary source of individual variability is speakers' prior experience, linguistic or otherwise (e.g., Baker and Trofimovich, 2005; Bohn and Flege, 1997; Flege et al., 1997, 1996; Gorba and Cebrian, 2021; Lev-Ari, 2018; Lev-Ari and Peperkamp, 2016). For instance, Flege et al. (1997) assessed the effect of experience with English on non-native speakers' (including L1 German, L1 Spanish, L1 Mandarin, and L1 Korean speakers) production and perception of English vowels. They found that experienced non-native participants (i.e., who resided in the US for around 7.3 years) produced and perceived English vowels more accurately than relatively inexperienced non-native participants (i.e., who resided in the US for around .7 years). Lev-Ari and Peperkamp (2016) observed that individuals from communities where there is a higher proportion of non-trill language speakers are less likely to misperceive a foreign tap as a trill, and concluded that individuals' environment can influence their perception by shaping their linguistic expectations. In addition, Lev-Ari (2018) found that people with larger social networks are better at vowel perception in noise, indicating that exposure to multiple speakers, and therefore to variable speech patterns, facilitates phonological performance.

Extrapolating from the findings just reviewed, it is expected that the more experience one has in engaging with English, the more native-like one is in terms of the use of post-stop F0 for voicing. Statistically, it is predicted that, in the English context, those with more experience would show a higher weight for post-stop F0 in both production and perception. Here the L1 Mandarin-L2 English bilinguals' language experience with English is approximated by *summative language experience* (SLE), which was invented by the author and is based on the participant's language history:

- Age of acquisition for English (since birth, 1, 2, ..., 20);
- Age at which you became comfortable using English (as early as I can re-

member, not yet, 1, 2, ..., 20);

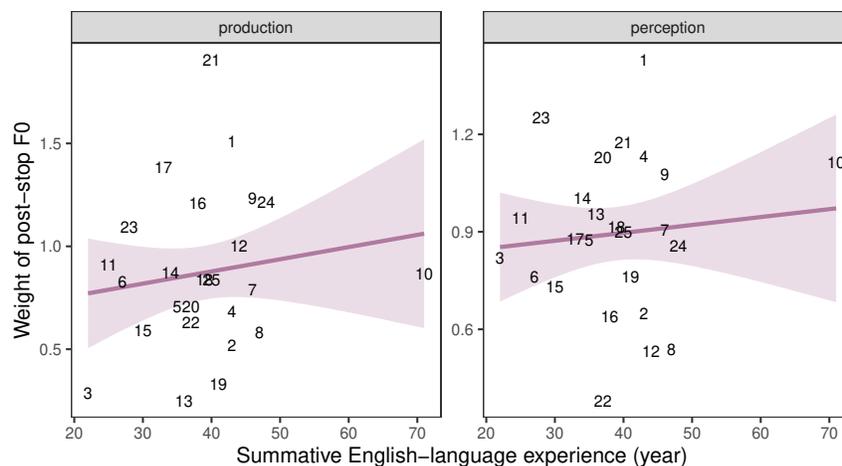
- Years of schooling in English (0, 1, 2, ..., 20);
- Years spent in a country or region where English is spoken (0, 1, 2, ..., 20);
- Years spent in a family where English is spoken (0, 1, 2, ..., 20);
- Years spent in a work or school environment where English is spoken (0, 1, 2, ..., 20).

The SLE score is then the sum of the numerical value given to each point, with the following exceptions: the first two items are scored in the reverse (i.e., a “20” response is worth 0, a “19” is worth 1, and so on), phrasal responses, such as “since birth” and “for as long as I can remember” for the first two questions, are worth 20, and “not yet” is worth 0. A higher SLE score hence means more overall experience with English.

Figure 6.2 plots individual mean post-stop F0 weights as a function of individual SLE scores in both production and perception. As is clear from the figure, a simple linear model does not provide evidence for a robust correlation between language experience and the post-stop F0 weight on an individual level in either modality (production:  $\bar{\beta}_{\text{SLE}} = .0057$ , 89% CrI =  $[-.0082, .0198]$ ,  $p(\beta_{\text{SLE}} > 0) = .75$ ; perception:  $\bar{\beta}_{\text{SLE}} = .0025$ , 89% CrI =  $[-.0066, .0115]$ ,  $p(\beta_{\text{SLE}} > 0) = .67$ ) despite the appearance of a weak positive trend. The prediction that bilinguals with more experience with English will show higher weights for post-stop F0 is not supported.

### **Language Dominance**

Another source of individual variability that is related to language experience but is distinct from it is language dominance: “the degree of bilingualism manifested by individuals who know two languages, that is, the relative level of proficiency in each of the languages” (Hemàndez-Chàvez et al., 1978, p. 41). For example, Hazan and Boulakia (1993) investigated the effect of language dominance on the use of VOT and spectral information in producing and perceiving English and French



**Figure 6.2:** L1 Mandarin-L2 English bilinguals’ production and perceptual weights for post-stop F0 in the English context as a function of individual summative language experience (SLE) scores. The lines represent linear regression results of post-stop F0 weight against SLE score. The shaded areas cover the 89% confidence interval of the regression. Each number represents a participant, and the same participant is represented by the same number in the two panels.

bilabial stops. They compared the performance from four groups: English monolinguals, English dominant English-French bilinguals, French dominant English-French bilinguals, and French monolinguals. In their production of the /pɛn/-/bɛn/ minimal pair in both languages (i.e., *pen* vs. *Ben* in English, and *peine* vs. *benne* in French), bilinguals shifted their VOTs according to the language context but did not always produce monolingual-like VOTs in their non-dominant language. In perception, the English-dominant bilinguals showed a greater use of spectral changes at vowel onset than the French-dominant bilinguals, regardless of the language context.

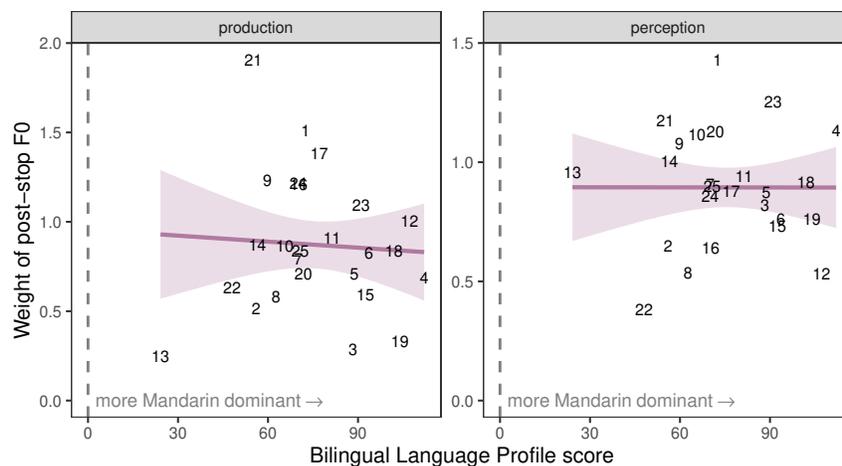
Casillas (2015) examined the production and perception of the English /i/-/ɪ/ contrast in three groups—native English speakers (NE), L2-dominant early learners of English who are no longer proficient in their L1 Spanish (EL), and L1-Spanish late-onset learners of English (LL). The production experiment revealed that the ELs did not differ from the NEs with regard to the duration and spectral

centroids of the contrast. The results from the perception experiment showed that both the NE and EL groups weighted spectral information more heavily than duration, though the ELs gave relatively more importance to duration than the NEs.

Similarly, Casillas and Simonet (2016) studied these three groups' production and perception of the /æ/-/a/ contrast in English. The production data revealed a clear difference between the NE group and the two learner groups: the NEs displayed two non-overlapping acoustic distributions, but the two learner groups merged the /æ/-/a/ contrast to some extent, though the LL group showed more extreme overlap for the contrast than the EL group. The perceptual identification experiment found that the two learner groups diverged from the NE group in the use of spectral cues in a scalar progression, with the LL group differing from the NE group more than the EL group.

Overall, these findings suggest that the more dominant one is in the target language on a bilingual spectrum, the more native-like one is in using the cues defining the sound categories in the target language. Applying this argument to the current study, it is expected that the more dominant one is in English, the more native-like one is in terms of the use of post-stop F0 for voicing in the English task, that is, having a higher weight for post-stop F0. Here, the L1 Mandarin-L2 English bilinguals' language dominance is characterized by the Bilingual Language Profile (BLP) score (Birdsong et al., 2012), which ranges from  $-218$  to  $+218$ . A score near 0 indicates balanced bilingualism, and, in the current research setting, more positive scores mean more Mandarin dominance, and more negative scores more English dominance.

Individual mean post-stop F0 weights as a function of individual BLP scores in both production and perception are depicted in Figure 6.3. As can be seen from the figure, simple linear regression does not indicate any discernible patterns in either modality (production:  $\bar{\beta}_{\text{BLP}} = -.0012$ , 89% CrI =  $[-.0079, .0056]$ ,  $p(\beta_{\text{BLP}} < 0) = .61$ ; perception:  $\bar{\beta}_{\text{BLP}} = -.000046$ , 89% CrI =  $[-.0043, .0042]$ ,  $p(\beta_{\text{BLP}} < 0) = .51$ ). The prediction that more dominance in English is correlated with more reliance on post-stop F0 is not borne out. Therefore, it seems that language dominance does not account for the observed individual differences, at least in the current research context with the metrics adopted.



**Figure 6.3:** L1 Mandarin-L2 English bilinguals’ production and perceptual weights for post-stop F0 in the English context as a function of individual Bilingual Language Profile (BLP) scores. The lines represent linear regression results of post-stop F0 weight against BLP score. The shaded areas cover the 89% confidence interval of the regression. Each number represents a participant, and the same participant is represented by the same number in the two panels.

### Production-Perception Interface

Individual differences in cue weighting also raise questions about the nature of the link between speech production and perception. Despite the general finding that perceptual patterns reflect the production norms on a population level, the correlation between individuals’ relative use of the post-stop F0 cue in production and perception is fairly weak at best. The failure to find a production-perception link on an individual level echoes previous studies examining this relationship in both L1 speech (e.g., Idemaru et al., 2012; Schertz et al., 2020, 2015; Shultz et al., 2012) and L2 speech (e.g., Bohn and Flege, 1997). As discussed in Section 3.5.5, such a lack of strong association between production and perception calls into questions gestural theories, such as the Motor Theory (Lieberman and Mattingly, 1985) and the Direct Realism (Fowler, 1986), that view speech perception as guided by the recovery of gestures in the underlying signal.

Lack of a direct production-perception connection also highlights the potential

influence of the difference in goals across modalities (Shultz et al., 2012; Yu and Zellou, 2019). The nature of the decision a listener makes and how representational memory is accessed across various perceptual tasks can help explain some of the conflicting empirical results in the literature. Zellou (2017), for instance, found that, in a task where listeners had to make explicit judgments about nasality in context (i.e., listeners were instructed to indicate which item of the pair contained a vowel that sounded more nasal), their responses did not correlate with their production; however, she did observe a significant correlation when the task required the discrimination of vocalic nasality in different contexts (i.e., each trial contains two pairs of stimuli, with one pair containing acoustically identical vowels and the other pair containing acoustically different vowels; listeners were instructed to determine which pair of words contained vowels that sounded most different). She interpreted this finding as suggesting that the nature of the perceptual task could affect whether or not listeners resort to their idiosyncratic production repertoires, as opposed to expectations based on experience accrued across a speech community.

In sum, further investigation is required to understand the conditions under which a production-perception link emerges. The literature has suggested that methodological issues are at the core of this research program. For instance, Schertz and Clare (2019) point out that the acoustic spaces that standard production and perception experiments tap into are often very different. That is, standard production tasks permit speakers to produce cues in an unconstrained way, whereas perception tasks often have stimuli that are unlikely to arise from natural production. As such, perceptual weights might be driven by tokens that are very rarely observed in production, which could in turn result in a lack of apparent link between the two modalities. How to address these methodological asymmetries and reconcile the apparently contradicting results can therefore serve as an active force in advancing our knowledge in both production and perception.

## **6.5 Conclusion**

The corpus study and the set of parallel experimental studies of post-stop F0 in cueing stop voicing across languages and modalities undertaken for this work offer insight into between- and within-language variation in the ways speakers and lis-

teners use post-stop F0 to define sound categories, as well as the process by which listeners vary cue weights in response to additional cognitive load.

The results from the corpus study indicate that, in Mandarin, the post-stop F0 next to an aspirated stop is *lower* than that next to an unaspirated one. This corpus study also speaks to the importance of F0 normalization methods. The production and perception experiments provide evidence that L1 Mandarin-L2 English bilinguals use post-stop F0 as a cue for stop voicing in their L1 and L2 production and perception. The experiment-based production data suggests that the bilinguals tend to produce aspirated/voiceless tokens with a *higher* post-stop F0 than unaspirated/voiced tokens, which runs counter to the pattern revealed in the corpus study. These conflicting outcomes might be attributable to L2-to-L1 cross-linguistic influence on the part of the bilinguals. The perceptual identification data indicates that, all else being equal, a higher post-stop F0 gives rise to more aspirated/voiceless responses than a lower post-stop F0. Furthermore, their use of post-stop F0 in both production and perception is flexible and sensitive to the current linguistic context, as reflected in the change of reliance on the post-stop F0 weight. This flexibility, nonetheless, is still constrained by L1 usage patterns in perception. The use of post-stop F0 in perception is also modulated by additional cognitive load, which chiefly enlarges the variability of cue weights. A cross-linguistic comparison of post-stop F0 difference in production is consistent with the view that post-stop F0 perturbation has its origin in an automatic process but is also subject to (subconscious) control.

This work also explores the interface between production and perception on an individual level. Despite the fact that there are robust individual variability in both production and perception, the individual differences in perception are at best weakly correlated with individual production patterns.

Overall, this work underscores the complex interplay between learned, language-specific knowledge, cognitive load, and methodological considerations in speakers' and listeners' use of post-stop F0 define the voicing contrast in stops. All the data and scripts for analyses can be found at [https://osf.io/jxuw7/?view\\_only=23ad137a075244d9bf5df99789a20cc9](https://osf.io/jxuw7/?view_only=23ad137a075244d9bf5df99789a20cc9).

# Bibliography

- Abramson, A. S. and Lisker, L. (1985). Relative power of cues: F0 shift versus voice timing. In *Phonetic linguistics: Essays in honor of Peter Ladefoged*, ed. V. A. Fromkin (New York, NY: Academic Press). 25–33 → pages 2, 56
- Abramson, A. S. and Whalen, D. H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics* 63, 75–86. doi:10.1016/j.wocn.2017.05.002 → page 11
- Allen, J. S. and Miller, J. L. (1999). Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. *The Journal of the Acoustical Society of America* 106, 2031–2039. doi:10.1121/1.427949 → page 91
- Amengual, M. (2018). Asymmetrical interlingual influence in the production of Spanish and English laterals as a result of competing activation in bilingual language processing. *Journal of Phonetics* 69, 12–28. doi:10.1016/j.wocn.2018.04.002 → page 192
- Amengual, M. (2021). The acoustic realization of language-specific phonological categories despite dynamic cross-linguistic influence in bilingual and trilingual speech. *The Journal of the Acoustical Society of America* 149, 1271–1284. doi:10.1121/10.0003559 → pages 3, 62, 112
- Antoniou, M., Best, C. T., Tyler, M. D., and Kroos, C. (2010). Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2. *Journal of Phonetics* 38, 640–653. doi:10.1016/j.wocn.2010.09.005 → page 112
- Antoniou, M., Tyler, M. D., and Best, C. T. (2012). Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of Phonetics* 40, 582–594. doi:10.1016/j.wocn.2012.05.005 → page 112

- Baker, W. and Trofimovich, P. (2005). Interaction of native- and second-language vowel system(s) in early and late bilinguals. *Language and Speech* 48, 1–27. doi:10.1177/00238309050480010101 → pages 196, 197
- Barlow, J. A. (2014). Age of acquisition and allophony in Spanish-English bilinguals. *Frontiers in Psychology* 5, 1–14. doi:10.3389/fpsyg.2014.00288 → page 192
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68, 255–278. doi:10.1016/j.jml.2012.11.001 → pages 35, 74
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67, 1–48. doi:10.18637/jss.v067.i01 → page 34
- Beddor, P. S. (2009). A coarticulatory path to sound change. *Language* 85, 785–821. doi:10.1353/lan.0.0165 → page 157
- Beddor, P. S. (2015). The relation between language users' perception and production repertoires. In *Proceedings of the 18th International Congress of Phonetic Sciences* (Glasgow: University of Glasgow), 1–9 → page 157
- Beddor, P. S., Coetzee, A. W., Styler, W., McGowan, K. B., and Boland, J. E. (2018). The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. *Language* 94, 931–968. doi:10.1353/lan.2018.0051 → page 115
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In *Speech perception and linguistic experience: Issues in cross-language research*, ed. W. Strange (Timonium, MD: York Press), chap. 6. 171–204 → page 119
- Best, C. T. and Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In *Language experience in second language speech learning: In honor of James Emil Flege*, eds. O.-S. Bohn and M. J. Munro (Amsterdam: John Benjamins Publishing Company). 13–34. doi:10.1075/llt.17.07bes → pages 62, 118, 119
- Bijarsari, S. E. (2021). A current view on dual-task paradigms and their limitations to capture cognitive load. *Frontiers in Psychology* 12, 1–6. doi:10.3389/fpsyg.2021.648586 → page 158

- Birdsong, D., Gertken, L. M., and Amengual, M. (2012). Bilingual Language Profile: An easy-to-use instrument to assess bilingualism. [Web page: <https://sites.la.utexas.edu/bilingual/>] → page 200
- Blicher, D. L., Diehl, R. L., and Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone2/Tone3 distinction: Evidence of auditory enhancement. *Journal of Phonetics* 18, 37–49. doi:10.1016/S0095-4470(19)30357-2 → pages 55, 90
- Boersma, P. and Weenink, D. (2021). Praat: Doing phonetics by computer. [Computer program: <https://www.fon.hum.uva.nl/praat/>] → pages 14, 15, 17, 24, 69, 89
- Bohn, O.-S. and Flege, J. E. (1997). Perception and production of a new vowel category by adult second language learners. In *Second-language speech: Structure and process*, eds. A. James and J. Leather (Berlin: De Gruyter Mouton). 53–73. doi:10.1515/9783110882933.53 → pages 196, 197, 201
- Bolinger, D. L. (1961). Contrastive accent and contrastive stress. *Language* 37, 83–96. doi:10.2307/411252 → page 2
- Bullock, B. E. and Toribio, A. J. (2009). Trying to hit a moving target: On the sociophonetics of code-switching. In *Multidisciplinary approaches to code switching*, eds. L. Isurin, D. Winford, and K. de Bot (John Benjamins Publishing Company), Studies in bilingualism, chap. 8. 189–206. doi:10.1075/sibil.41.12bul → page 192
- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80, 1–28. doi:10.18637/jss.v080.i01 → pages 34, 74, 129
- Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal* 10, 395–411. doi:10.32614/RJ-2018-017 → page 34
- Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software* 100, 1–54. doi:10.18637/jss.v100.i05 → page 34
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., et al. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software* 76, 1–32. doi:10.18637/jss.v076.i01 → page 72

- Casillas, J. V. (2015). Production and perception of the /i/-/ɪ/ vowel contrast: The case of L2-dominant early learners of English. *Phonetica* 72, 182–205. doi:10.1159/000431101 → pages 197, 199
- Casillas, J. V. and Simonet, M. (2016). Production and perception of the english /æ/-/ɑ/ contrast in switched-dominance speakers. *Second Language Research* 32, 171–195. doi:10.1177/0267658315608912 → pages 197, 200
- Casillas, J. V. and Simonet, M. (2018). Perceptual categorization and bilingual language modes: Assessing the *double phonemic boundary* in early and late bilinguals. *Journal of Phonetics* 71, 51–64. doi:10.1016/j.wocn.2018.07.002 → pages 62, 112
- Chen, L., Chao, K.-Y., Peng, J.-F., and Yang, J.-C. (2008). A cross-language study of stop aspiration: English and Mandarin Chinese. In *2008 Tenth IEEE International Symposium on Multimedia*. 556–561. doi:10.1109/ISM.2008.86 → page 125
- Chen, Y. (2011). How does phonology guide phonetics in segment-*f0* interaction? *Journal of Phonetics* 39, 612–625. doi:10.1016/j.wocn.2011.04.001 → pages 3, 14, 15, 16, 17, 18, 20, 24, 25, 51, 53, 56, 58
- Chodroff, E. and Baese-Berk, M. (2019). Constraints on variability in the voice onset time of L2 English stop consonants. In *Proceedings of the 19th International Congress of Phonetic Sciences*, eds. S. Calhoun, P. Escudero, M. Tabain, and P. Warren (Melbourne), 661–665 → page 85
- Choi, W., Tong, X., and Samuel, A. G. (2019). Better than native: Tone language experience enhances English lexical stress discrimination in Cantonese-English bilingual listeners. *Cognition* 189, 188–192. doi:10.1016/j.cognition.2019.04.004 → page 120
- Ciaccio, L. A. and Veríssimo, J. (2020). Investigating variability in morphological processing with Bayesian distributional models. [PsyArXiv: <https://psyarxiv.com/qk5cu/>]. doi:10.31234/osf.io/qk5cu → page 181
- Clayards, M. (2018). Individual talker and token covariation in the production of multiple cues to stop voicing. *Phonetica* 75, 1–23. doi:10.1159/000448809 → pages 7, 23, 69, 77
- Clayards, M., Tanenhaus, M. K., Aslin, R. N., and Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition* 108, 804–809. doi:10.1016/j.cognition.2008.04.004 → page 4

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Mahwah, NJ: Lawrence Erlbaum Associates), 2 edn. → page 77
- Cole, J., Kim, H., Choi, H., and Hasegawa-Johnson, M. (2007). Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech. *Journal of Phonetics* 35, 180–209. doi:10.1016/j.wocn.2006.03.004 → page 69
- Connell, B. (2001). Downdrift, downstep, and declination. In *Typology of African prosodic systems workshop*. 1–8 → pages 36, 48
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology* 6, 84–107. doi:10.1016/0010-0285(74)90005-X → page 4
- Dale, R., Kehoe, C., and Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition* 35, 15–28. doi:10.3758/BF03195938 → page 96
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods* 47, 1–12. doi:10.3758/s13428-014-0458-y → pages 68, 96
- Ding, H., Zhan, Y., Yuan, J., and Liao, S. (2018). Production of English stops by Mandarin Chinese learners. In *Proceedings of Speech Prosody 2018*. 888–892. doi:10.21437/SpeechProsody.2018-179 → page 125
- Disner, S. F. (1980). Evaluation of vowel normalization procedures. *The Journal of the Acoustical Society of America* 67, 253–261. doi:10.1121/1.383734 → page 22
- Dmitrieva, O. (2019). Transferring perceptual cue-weighting from second language into first language: Cues to voicing in Russian speakers of English. *Journal of Phonetics* 73, 128–143. doi:10.1016/j.wocn.2018.12.008 → page 197
- Dmitrieva, O., Llanos, F., Shultz, A. A., and Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset  $f_0$  as a secondary voicing cue in Spanish and English. *Journal of Phonetics* 49, 77–95. doi:10.1016/j.wocn.2014.12.005 → pages 11, 14, 21, 23, 56, 69
- Ewan, W. (1976). *Laryngeal behavior in speech*. Ph.D. thesis, University of California, Berkeley, Berkeley, CA → pages 19, 56

- Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics* 15, 47–65. doi:10.1016/S0095-4470(19)30537-6 → page 192
- Flege, J. E. (1995). Second language speech learning theory, findings, and problems. In *Speech perception and linguistic experience: Issues in cross-language research*, ed. W. Strange (Timonium, MD: York Press), chap. 8. 233–277 → pages 118, 191
- Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In *Laboratory Phonology 9*, eds. J. Cole and J. I. Hualde (Berlin: Mouton de Gruyter). 353–381 → pages 118, 191
- Flege, J. E. and Bohn, O.-S. (2021). The revised speech learning model (SLM-r). In *Second language speech learning: Theoretical and empirical progress*, ed. R. Wayland (Cambridge: Cambridge University Press), chap. 1. 3–83. doi:10.1017/9781108886901.002 → pages 61, 151, 191
- Flege, J. E., Bohn, O.-S., and Jang, S. (1997). Effects of experience on non-native speakers’ production and perception of English vowels. *Journal of Phonetics* 25, 437–470. doi:10.1006/jpho.1997.0052 → pages 151, 196, 197
- Flege, J. E. and Eefting, W. (1987a). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication* 6, 185–202. doi:10.1016/0167-6393(87)90025-2 → page 151
- Flege, J. E. and Eefting, W. (1987b). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics* 15, 67–83. doi:10.1016/S0095-4470(19)30538-8 → page 151
- Flege, J. E., Takagi, N., and Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults’ perception of /ɹ/ and ///. *The Journal of the Acoustical Society of America* 99, 1161–1173. doi:10.1121/1.414884 → page 197
- Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics* 14, 3–28. doi:10.1016/S0095-4470(19)30607-2 → pages 7, 59, 115, 119, 151, 197, 201
- Francis, A. L., Ciocca, V., Wong, V. K. M., and Chan, J. K. L. (2006). Is fundamental frequency a cue to aspiration in initial stops? *The Journal of the Acoustical Society of America* 120, 2884–2895. doi:10.1121/1.2346131 → pages 51, 56, 57, 113, 192

- Francis, A. L., Kaganovich, N., and Driscoll-Huber, C. (2008). Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *The Journal of the Acoustical Society of America* 124, 1234–1251. doi:10.1121/1.2945161 → pages 59, 157
- Fricke, M., Kroll, J. F., and Dussias, P. E. (2016). Phonetic variation in bilingual speech: A lens for studying the production-comprehension link. *Journal of Memory and Language* 89, 110–137. doi:10.1016/j.jml.2015.10.001 → page 192
- Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech* 1, 126–152. doi:10.1177/002383095800100207 → page 2
- Fulop, S. A. and Scott, H. J. M. (2021). Consonant voicing in the Buckeye corpus. *The Journal of the Acoustical Society of America* 149, 4190–4197. doi:10.1121/10.0005199 → page 91
- Gabry, J. and Češnovar, R. (2021). *cmdstanr: R interface to ‘CmdStan’*. <https://mc-stan.org/cmdstanr>, <https://discourse.mc-stan.org> → pages 72, 99
- Gandour, J. (1974). Consonant types and tone in siamese. *Journal of Phonetics* 2, 337–350. doi:10.1016/S0095-4470(19)31303-8 → pages 51, 56, 57, 113
- Gandour, J. T. (1978). The perception of tone. In *Tone: A linguistic survey*, ed. V. A. Fromkin (New York, NY: Academic Press), chap. 2. 41–76 → pages 2, 56
- Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6, 110–125. doi:10.1037/0096-1523.6.1.110 → page 157
- Gao, J. and Arai, T. (2018). F0 perturbation in a “pitch-accent” language. In *Proceedings of the Sixth International Symposium on Tonal Aspects of Languages*. 56–60. doi:10.21437/TAL.2018-12 → pages 14, 56
- Garner, W. R. (1974). *The processing of information and structure* (Mahwah, NJ: Lawrence Erlbaum Associates). doi:10.4324/9781315802862 → page 121
- Garner, W. R. (1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology* 8, 98–123. doi:10.1016/0010-0285(76)90006-2 → page 121
- Garner, W. R. and Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology* 1, 225–241. doi:10.1016/0010-0285(70)90016-2 → page 121

- Gonzales, K., Byers-Heinlein, K., and Lotto, A. J. (2019). How bilinguals perceive speech depends on which language they think they're hearing. *Cognition* 182, 318–330. doi:10.1016/j.cognition.2018.08.021 → page 112
- Gonzales, K. and Lotto, A. J. (2013). A Bafri, un Pafri: Bilinguals' pseudoword identifications support language-specific phonetic systems. *Psychological Science* 24, 2135–2142. doi:10.1177/0956797613486485 → page 112
- Gorba, C. and Cebrian, J. (2021). The role of L2 experience in L1 and L2 perception and production of voiceless stops by English learners of Spanish. *Journal of Phonetics* 88, 1–25. doi:10.1016/j.wocn.2021.101094 → page 197
- Gordon, P. C., Eberhardt, J. L., and Rueckl, J. G. (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology* 25, 1–42. doi:10.1006/cogp.1993.1001 → pages 154, 155, 156, 157, 158, 159, 160, 177, 181, 183, 184, 185, 189
- Grage, T., Schoemann, M., Kieslich, P. J., and Scherbaum, S. (2019). Lost to translation: How design factors of the mouse-tracking procedure impact the inference from action to cognition. *Attention, Perception, & Psychophysics* 81, 2538–2557. doi:10.3758/s13414-019-01889-z → page 5
- Guion, S. G. (2003). The vowel systems of Quichua-Spanish bilinguals. *Phonetica* 60, 98–128. doi:10.1159/000071449 → page 192
- Guo, Y. (2020). *Production and perception of laryngeal contrasts in Mandarin and English by Mandarin speakers*. Ph.D. thesis, George Mason University, Fairfax, VA → pages 3, 12, 14, 15, 16, 17, 18, 20, 22, 25, 47, 53, 56, 57, 58, 61, 66, 86, 92, 109, 112, 139
- Halle, M. and Stevens, K. N. (1971). A note on laryngeal features. *MIT Research Laboratory of Electronics Quaterly Progress Report* 101, 198–214. doi:10.1515/9783110871258.45 → pages 7, 19, 195
- Han, M. S. and Weitzman, R. S. (1970). Acoustic features of Korean /P, T, K/, /p, t, k/ and /p<sup>h</sup>, t<sup>h</sup>, k<sup>h</sup>/. *Phonetica* 22, 112–128. doi:10.1159/000259311 → page 56
- Hanson, H. M. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in English. *The Journal of the Acoustical Society of America* 1, 425–441. doi:10.1121/1.3021306 → pages 2, 11, 13, 56, 112

- Hazan, V. L. and Boulakia, G. (1993). Perception and production of a voicing contrast by French-English bilinguals. *Language and Speech* 36, 17–38. doi:10.1177/002383099303600102 → pages 197, 198
- Hemàndez-Chàvez, E., Burt, M., and Dulay, H. (1978). Language dominance and proficiency testing: Some general considerations. *NABE Journal* 3, 41–54. doi:10.1080/08855072.1978.10668343 → page 198
- Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America* 97, 3099–3111. doi:10.1121/1.411872 → page 88
- Ho, A. T. (1976). The acoustic variation of Mandarin tones. *Phonetica* 33, 353–367. doi:10.1159/000259792 → page 55
- Holt, L. L. and Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America* 119, 3059–3071. doi:10.1121/1.2188377 → page 59
- Hombert, J.-M. (1978). Consonant types, vowel quality, and tone. In *Tone: A linguistic survey*, ed. V. A. Fromkin (New York, NY: Academic Press), chap. 3. 77–111 → pages 2, 56, 57, 113
- Hombert, J.-M., Ohala, J. J., and Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language* 55, 37–58. doi:10.2307/412518 → pages 2, 7, 19, 56, 195
- Honda, K., Hirai, H., Masaki, S., and Shimada, Y. (1999). Role of vertical larynx movement and cervical lordosis in F0 control. *Language and Speech* 42, 401–411. doi:10.1177/00238309990420040301 → page 19
- Hoole, P. and Honda, K. (2011). Automaticity vs. feature-enhancement in the control of segmental F0. In *Where do phonological features come from? Cognitive, physical and developmental bases of distinctive speech categories*, eds. G. N. Clements and R. Ridouane (Amsterdam: John Benjamins Publishing Company). 131–172. doi:10.1075/lfab.6.06hoo → page 21
- House, A. S. and Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America* 25, 105–113. doi:10.1121/1.1906982 → pages 2, 12, 56

- Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones* (Cambridge: Cambridge University Press) → pages 3, 13, 14, 15, 16, 17, 18, 20, 24, 56, 58
- Huang, S., Liu, J., Wu, X., Wu, L., Yan, Y., and Qin, Z. (1997). Mandarin broadcast news speech and transcripts. [Corpus data: <https://catalog ldc.upenn.edu/LDC98T24>] → page 26
- Idemaru, K., Holt, L. L., and Seltman, H. (2012). Individual differences in cue weights are stable across time: The case of Japanese stop lengths. *The Journal of the Acoustical Society of America* 132, 3950–3964. doi:10.1121/1.4765076 → page 201
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, B47–B57. doi:10.1016/S0010-0277(02)00198-1 → pages 3, 60
- Jessen, M. and Roux, J. C. (2002). Voice quality differences associated with stops and clicks in Xhosa. *Journal of Phonetics* 30, 1–52. doi:10.1006/jpho.2001.0150 → page 56
- Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In *Talker variability in speech processing*, eds. K. Johnson and J. W. Mullennix (San Diego, CA: Academic Press), chap. 8. 145–165 → page 59
- Johnson, K. (2020). The  $\Delta F$  method of vocal tract length normalization for vowels. *Laboratory Phonology* 11, 1–16. doi:10.5334/labphon.196 → pages 22, 29
- Johnson, K. A. (2021). Leveraging the uniformity framework to examine crosslinguistic similarity for long-lag stops in spontaneous Cantonese-English bilingual speech. In *Proceedings of Interspeech 2021*. 2671–2675. doi:10.21437/Interspeech.2021-1780 → page 85
- Jongman, A. and Wade, T. (2007). Acoustic variability and perceptual learning: The case of non-native accented speech. In *Language experience in second language speech learning: In honor of James Emil Flege*, eds. O.-S. Bohn and M. J. Munro (Amsterdam: John Benjamins Publishing Company), chap. 8. 135–150. doi:10.1075/llt.17.14jon → page 148

- Jun, S.-A. (1996). Influence of microprosody on macroprosody: A case of phrase initial strengthening. *UCLA Working Papers in Phonetics* 92, 97–116 → page 56
- Kang, K.-H. and Guion, S. G. (2008). Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *The Journal of the Acoustical Society of America* 124, 3909–3917. doi:10.1121/1.2988292 → page 156
- Kang, Y. (2014). Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics* 45, 76–90. doi:10.1016/j.wocn.2014.03.005 → page 156
- Kartushina, N. and Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in Psychology* 5, 1–17. doi:10.3389/fpsyg.2014.01246 → page 150
- Kataoka, R. (2011). *Phonetic and cognitive bases of sound change*. Ph.D. thesis, University of California, Berkeley, Berkeley, CA → page 59
- Keating, P. and Kuo, G. (2012). Comparison of speaking fundamental frequency in English and Mandarin. *The Journal of the Acoustical Society of America* 132, 1050–1060. doi:10.1121/1.4730893 → page 81
- Keyser, S. J. and Stevens, K. N. (2006). Enhancement and overlap in the speech chain. *Language* 82, 33–63 → page 196
- Kieslich, P. J., Henninger, F., Wulff, D. U., Haslbeck, J. M. B., and Schulte-Mecklenbeck, M. (2019). Mouse-tracking: A practical guide to implementation and analysis. In *A handbook of process tracing methods*, eds. M. Schulte-Mecklenbeck, A. Kühberger, and J. G. Johnson (New York, NY: Routledge), chap. 8. 2 edn., 111–130. doi:10.4324/9781315160559 → page 5
- Kieslich, P. J., Schoemann, M., Grage, T., Hepp, J., and Scherbaum, S. (2020). Design factors in mouse-tracking: What makes a difference? *Behavior Research Methods* 52, 317–341. doi:10.3758/s13428-019-01228-y → page 5
- Kim, H. and Tremblay, A. (2021). Korean listeners' processing of suprasegmental lexical contrasts in Korean and English: A cue-based transfer approach. *Journal of Phonetics* 87, 1–15. doi:10.1016/j.wocn.2021.101059 → page 120
- Kim, M.-R. (2013). VOT merger and f0 maximization between the lax and aspirated stop in sound change. *The Linguistic Association of Korean Journal* 21, 1–20. doi:10.24303/lakdoi.2013.21.2.1 → page 156

- Kingston, J. (2007). Segmental influences on F<sub>0</sub>: Automatic or controlled. In *Tones and Tunes: Volume 2: Experimental studies in word and sentence prosody*, eds. C. Gussenhoven and T. Riad (Berlin: De Gruyter Mouton). 171–210. doi:10.1515/9783110207576.2.171 → pages 21, 196
- Kingston, J. and Diehl, R. L. (1994). Phonetic knowledge. *Language* 70, 419–454. doi:10.1353/lan.1994.0023 → pages 7, 11, 20, 21, 56, 57, 63, 65, 111, 195, 196
- Kingston, J., Diehl, R. L., Kirk, C. J., and Castleman, W. A. (2008). On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics* 36, 28–54. doi:10.1016/j.wocn.2007.02.001 → pages 7, 57
- Kirby, J. P. (2018). Onset pitch perturbations and the cross-linguistic implementation of voicing: Evidence from tonal and non-tonal languages. *Journal of Phonetics* 71, 326–354. doi:10.1016/j.wocn.2018.09.009 → pages 12, 112, 114
- Kirby, J. P. and Ladd, D. R. (2016). Effects of obstruent voicing on vowel F<sub>0</sub>: Evidence from “true voicing” languages. *The Journal of the Acoustical Society of America* 140, 2400–2411. doi:10.1121/1.4962445 → pages 12, 13, 51, 56, 112
- Kohler, K. J. (1982). F<sub>0</sub> in the production of lenis and fortis plosives. *Phonetica* 39, 199–218. doi:10.1159/000261663 → page 56
- Kohler, K. J. (1984). Phonetic explanation in phonology: The feature fortis/lenis. *Phonetica* 41, 150–174. doi:10.1159/000261721 → pages 7, 19, 20, 65, 111, 195
- Kong, E. J. and Lee, H. (2018). Attentional modulation and individual differences in explaining the changing role of fundamental frequency in Korean laryngeal stop perception. *Language and Speech* 61, 384–408. doi:10.1177/0023830917729840 → pages 154, 155, 156, 158, 159
- Koster, J. and McElreath, R. (2017). Multinomial analysis of behavior: Statistical methods. *Behavioral Ecology and Sociobiology* 71, 1–14. doi:10.1007/s00265-017-2363-8 → page 35
- Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (London: Academic Press), 2 edn. → page 99

- Ladefoged, P. (1963). Some physiological parameters in speech. *Language and Speech* 6, 109–119. doi:10.1177/002383096300600301 → page 20
- Ladefoged, P. (1967). *Three areas of experimental phonetics* (Oxford: Oxford University Press) → pages 7, 19, 20, 65, 111, 195
- Ladefoged, P. (1971). *Preliminaries to linguistic phonetics* (Chicago, IL: The University of Chicago Press) → page 19
- Ladefoged, P. (1974). Respiration, laryngeal activity and linguistics. In *Proceedings of the International Symposium on Ventilatory and Phonatory Control Systems*, ed. B. Wyke (Oxford: Oxford University Press), 299–314 → page 20
- Ladefoged, P. and Johnson, K. (2014). *A course in phonetics* (Stamford, CT: Cengage Learning), 7 edn. → page 2
- Ladefoged, P. and Maddieson, I. (1996). *The sounds of the world's languages* (Oxford: Blackwell Publishing Ltd) → page 120
- Lai, Y., Huff, C., Sereno, J., and Jongman, A. (2009). The raising effect of aspirated prevocalic consonants on  $f_0$  in Taiwanese. In *Proceedings of the 2nd International Conference on East Asian Linguistics*, eds. J. Brooke, G. Coppola, E. Görgülü, M. Mameni, E. Mileva, S. Morton, and A. Rimrott. 1–12 → page 51
- Lea, W. A. (1973). Segmental and suprasegmental influences on fundamental frequency contours. In *Consonant types & tone: Southern California occasional papers in linguistics no. 1*, ed. L. M. Hyman (Los Angeles, CA: The Linguistics Program, University of Southern California). 16–70 → pages 2, 12, 56
- Lee, B. and Sidtis, D. V. L. (2017). The bilingual voice: Vocal characteristics when speaking two languages across speech tasks. *Speech, Language and Hearing* 20, 174–185. doi:10.1080/2050571X.2016.1273572 → page 81
- Lee, H. and Jongman, A. (2012). Effects of tone on the three-way laryngeal distinction in Korean: An acoustic and aerodynamic comparison of the Seoul and South Kyungsang dialects. *Journal of the International Phonetic Association* 42, 145–169. doi:10.1017/S0025100312000035 → page 156
- Lee, L. and Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin

- Chinese. *Perception & Psychophysics* 53, 157–165. doi:10.3758/BF03211726  
→ pages 122, 150
- Lehiste, I. (1970). *Suprasegmentals* (Cambridge, MA: MIT Press) → page 2
- Lehiste, I. and Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. *The Journal of the Acoustical Society of America* 33, 419–425. doi:10.1121/1.1908681 → pages 2, 12, 56
- Lev-Ari, S. (2018). The influence of social network size on speech perception. *Quarterly Journal of Experimental Psychology* 71, 2249–2260. doi:10.1177/1747021817739865 → page 197
- Lev-Ari, S. and Peperkamp, S. (2016). How the demographic makeup of our community influences speech perception. *The Journal of the Acoustical Society of America* 139, 3076–3087. doi:10.1121/1.4950811 → page 197
- Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis* 100, 1989–2001. doi:10.1016/j.jmva.2009.04.008 → page 74
- Liberman, A. M., Delattre, P. C., Cooper, F. S., and Gerstman, L. J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs: General and Applied* 68, 1–13. doi:10.1037/h0093673 → page 55
- Liberman, A. M. and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition* 21, 1–36. doi:10.1016/0010-0277(85)90021-6 → pages 7, 59, 115, 151, 197, 201
- Liberman, M. (2014). Consonant effects on F0 in Chinese. [Blog post: <https://languagelog ldc.upenn.edu/nll/?p=12902>] → pages 22, 23, 24, 25, 27, 28, 47, 52
- Lindau, M. (1985). The story of /r/. In *Phonetic linguistics: Essays in honor of Peter Ladefoged*, ed. V. A. Fromkin (New York, NY: Academic Press). 157–168 → page 120
- Lisker, L. (1986). “Voicing” in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29, 3–11. doi:10.1177/002383098602900102 → page 54

- Löfqvist, A. (1975). Intrinsic and extrinsic  $F_0$  variations in Swedish tonal accents. *Phonetica* 31, 228–247. doi:10.1159/000259671 → pages 7, 19, 20
- Löfqvist, A., Baer, T., McGarr, N. S., and Story, R. S. (1989). The cricothyroid muscle in voicing control. *The Journal of the Acoustical Society of America* 85, 1314–1321. doi:10.1121/1.397462 → page 19
- Lotto, A. J., Sato, M., and Diehl, R. L. (2004). Mapping the task for the second language learner: The case of Japanese acquisition of /r/ and /l/. In *From sound to sense: 50+ years of discoveries in speech communication*. C–181–C–186 → page 60
- Luo, Q. (2018). *Consonantal effects on F0 in tonal languages*. Ph.D. thesis, Michigan State University, East Lansing, MI → pages 3, 12, 14, 15, 16, 17, 18, 20, 22, 24, 25, 47, 53, 56, 58
- Maldonado, M., Dunbar, E., and Chemla, E. (2018). Manipulated decision tasks to decode behavioral measures: The case of mouse-tracking. Manuscript → page 5
- Mattys, S. L., Barden, K., and Samuel, A. G. (2014). Extrinsic cognitive load impairs low-level speech perception. *Psychonomic Bulletin & Review* 21, 748–754. doi:10.3758/s13423-013-0544-7 → page 154
- Mattys, S. L. and Palmer, S. D. (2015). Divided attention disrupts perceptual encoding during speech recognition. *The Journal of the Acoustical Society of America* 137, 1464–1472. doi:10.1121/1.4913507 → page 154
- Mattys, S. L. and Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language* 65, 145–160. doi:10.1016/j.jml.2011.04.004 → pages 154, 155, 157, 158, 159, 160, 162, 164, 168, 177, 183, 185
- McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* (Boca Raton, FL: CRC Press) → page 35
- McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., and Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology: Human Perception and Performance* 34, 1609–1631. doi:10.1037/a0011747 → page 4
- McMurray, B., Tanenhaus, M. K., and Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition* 86, B33–B42. doi:10.1016/S0010-0277(02)00157-9 → page 4

- McMurray, B., Tanenhaus, M. K., and Aslin, R. N. (2009). Within-category VOT affects recovery from “lexical” garden-paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language* 60, 65–91. doi:10.1016/j.jml.2008.07.002 → page 4
- Mitterer, H. and Mattys, S. L. (2017). How does cognitive load influence speech perception? An encoding hypothesis. *Attention, Perception, & Psychophysics* 79, 344–351. doi:10.1016/j.jml.2011.04.004 → page 154
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., and Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics* 18, 331–340. doi:10.3758/BF03211209 → pages 3, 60
- Mohr, B. (1971). Intrinsic variations in the speech signal. *Phonetica* 23, 65–93. doi:10.1159/000259332 → page 56
- Morrison, G. S. and Kondaurova, M. V. (2009). Analysis of categorical response data: Use logistic regression rather than endpoint-difference scores or discriminant analysis. *The Journal of the Acoustical Society of America* 126, 2159–2162. doi:10.1121/1.3216917 → page 7
- Moulines, E. and Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication* 9, 453–467. doi:10.1016/0167-6393(90)90021-Z → page 90
- Nieuwenhuis, R., te Grotenhuis, M., and Pelzer, B. (2017). Weighted effect coding for observational data with wec. *The R Journal* 9, 477–485. doi:10.32614/RJ-2017-017 → pages 31, 34
- Ohala, J. J. (1978). Production of tone. In *Tone: A linguistic survey*, ed. V. A. Fromkin (New York, NY: Academic Press), chap. 1. 5–39 → pages 2, 20, 55
- Ohala, M. and Ohala, J. (1972). The problem of aspiration in Hindi phonetics. *Annual Bulletin, Research Institute of Logopedics and Phoniatics* 6, 39–46 → pages 7, 19, 20, 65, 111, 195
- Ohde, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *The Journal of the Acoustical Society of America* 75, 224–230. doi:10.1121/1.390399 → pages 2, 56
- Park, B. and Brünken, R. (2015). The rhythm method: A new method for measuring cognitive load—an experimental dual-task study. *Applied Cognitive Psychology* 29, 232–243. doi:10.1002/acp.3100 → page 184

- Peng, S. (1993). Cross-language influence on the production of Mandarin /f/ and /x/ and Taiwanese /h/ by native speakers of Taiwanese Amoy. *Phonetica* 50, 245–260. doi:10.1159/000261945 → page 192
- Phuong, V. T. (1981). *The acoustic and perceptual nature of tone in Vietnamese*. Ph.D. thesis, Australian National University, Canberra → pages 51, 57, 113
- Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In *Frequency and the emergence of linguistic structure*, eds. J. L. Bybee and P. J. Hopper (Amsterdam: John Benjamins Publishing Company). 137–157. doi:10.1075/tsl.45.08pie → page 59
- Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech* 46, 115–154. doi:10.1177/00238309030460020501 → page 59
- Ren, X. and Mok, P. (2021). Consonantal effects of aspiration on onset F0 in Cantonese. In *2021 12th International Symposium on Chinese Spoken Language Processing (ISCSLP)*. 1–5. doi:10.1109/ISCSLP49672.2021.9362106 → pages 56, 192
- Repp, B. H. and Lin, H.-B. (1990). Integration of segmental and tonal information in speech perception: A cross-linguistic study. *Journal of Phonetics* 18, 481–495. doi:10.1016/S0095-4470(19)30410-3 → pages 122, 150
- Rose, P. (2016). A comparison of normalisation strategies for citation tone F0 in four Chinese dialects. In *Proceedings of the 16th Australasian International Conference on Speech Science and Technology*, eds. C. Carignan and M. D. Tyler. 221–224 → pages 23, 73
- Schertz, J., Carbonell, K., and Lotto, A. J. (2020). Language specificity in phonetic cue weighting: Monolingual and bilingual perception of the stop voicing contrast in English and Spanish. *Phonetica* 77, 186–208. doi:10.1159/000497278 → pages 59, 88, 115, 139, 201
- Schertz, J., Cho, T., Lotto, A., and Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics* 52, 183–204. doi:10.1016/j.wocn.2015.07.003 → pages 60, 61, 77, 114, 201
- Schertz, J. and Clare, E. J. (2019). Phonetic cue weighting in perception and production. *WIREs Cognitive Science*, e1521 doi:10.1002/wcs.1521 → pages 151, 202

- Schertz, J. L. (2014). *The structure and plasticity of phonetic categories across languages and modalities*. Ph.D. thesis, The University of Arizona, Tucson, AZ  
→ page 114
- Schoemann, M., Lüken, M., Grage, T., Kieslich, P. J., and Scherbaum, S. (2019). Validating mouse-tracking: How design factors influence action dynamics in intertemporal decision making. *Behavior Research Methods* 51, 2356–2377. doi:10.3758/s13428-018-1179-4 → page 5
- Schoemann, M., O’Hora, D., Dale, R., and Scherbaum, S. (2021). Using mouse cursor tracking to investigate online cognition: Preserving methodological ingenuity while moving toward reproducible science. *Psychonomic Bulletin & Review* 28, 766–787. doi:10.3758/s13423-020-01851-3 → page 5
- Schultz, T. (2002). GlobalPhone: A multilingual speech and text database developed at Karlsruhe University. In *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP 2002)*. 345–348 → page 24
- Shimizu, K. (1994). F0 in phonation types of initial-stops. In *Proceedings of the 5th Australasian International Conference on Speech Science and Technology*. vol. 2, 650–655 → pages 57, 113
- Shook, A. and Marian, V. (2016). The influence of native-language tones on lexical access in the second language. *The Journal of the Acoustical Society of America* 139, 3102–3109. doi:10.1121/1.4953692 → page 123
- Shultz, A. A., Francis, A. L., and Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America* 132, EL95–EL101. doi:10.1121/1.4736711 → pages 23, 59, 77, 114, 201, 202
- Silva, D. J. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology* 23, 287–308. doi:10.1017/S0952675706000911 → page 156
- Sivula, T., Magnusson, M., and Vehtari, A. (2020). Uncertainty in Bayesian leave-one-out cross-validation based model comparison. doi:10.48550/arXiv.2008.10296 → page 36
- Slis, I. H. (1970). Articulatory measurements on voiced, voiceless and nasal consonants: A test of a model. *Phonetica* 21, 193–210. doi:10.1159/000259302 → pages 7, 19, 20

- Sonderegger, M., McAuliffe, M., and Bang, H.-Y. (2017). Segmental influences on F0: Cross-linguistic and interspeaker variability of precursors to sound change. In *4th Workshop on Sound Change: Accepted Abstracts*. 129–130 → page 24
- Spivey, M. J., Grosjean, M., and Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America* 102, 10393–10398. doi:10.1073/pnas.0503903102 → page 5
- Stan Development Team (2020). RStan: the R interface to Stan. [Software: <https://mc-stan.org/>]. R package version 2.21.2 → page 34
- Stoerber, M. (2019). Testing the effectiveness of mouse tracking in speech perception. Bachelor Thesis → page 5
- Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance* 7, 1074–1095. doi:10.1037/0096-1523.7.5.1074 → page 91
- Szakay, A. and Torgersen, E. (2019). A re-analysis of f0 in ethnic varieties of London English using REAPER. In *Proceedings of the 19th International Congress of Phonetic Sciences*, eds. S. Calhoun, P. Escudero, M. Tabain, and P. Warren (Melbourne), 1675–1678 → page 25
- Tanenhaus, M. K., Leiman, J. M., and Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior* 18, 427–440. doi:10.1016/S0022-5371(79)90237-8 → page 4
- Tanenhaus, M. K., Michael J. Spivey-Knowlton, K. M. E., and Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science* 268, 1632–1634. doi:10.1126/science.7777863 → page 4
- te Grotenhuis, M., Pelzer, B., Eisinga, R., Nieuwenhuis, R., Schmidt-Catran, A., and Konig, R. (2017a). A novel method for modelling interaction between categorical variables. *International Journal of Public Health* 62, 427–431. doi:10.1007/s00038-016-0902-0 → page 34
- te Grotenhuis, M., Pelzer, B., Nieuwenhuis, R., Schmidt-Catran, A., and Konig, R. (2017b). When size matters: Advantages of weighted effect coding in observational studies. *International Journal of Public Health* 62, 163–167. doi:10.1007/s00038-016-0901-1 → page 34

- Thomas, E. R. (2002). Instrumental phonetics. In *The handbook of language variation and change*, eds. J. K. Chambers, P. Trudgill, and N. Schilling-Estes (Malden, MA: Blackwell Publishing Ltd), chap. 7. 168–200 → page 22
- Thomas, E. R. (2011). *Sociophonetics: An introduction* (Basingstoke: Palgrave Macmillan) → page 22
- Tong, Y., Francis, A. L., and Gandour, J. T. (2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes* 23, 689–708. doi:10.1080/01690960701728261 → page 122
- Toscano, J. C. and McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science* 34, 434–464. doi:10.1111/j.1551-6709.2009.01077.x → pages 54, 59, 91
- Trofimovich, P. and John, P. (2011). When three equals tree: Examining the nature of phonological entries in L2 lexicons of Quebec speakers of English. In *Applying priming research to L2 teaching and learning: Insight from Psycholinguistics*, ed. P. Trofimovich (Amsterdam: John Benjamins Publishing Company), chap. 5. 105–129. doi:10.1075/llt.30.09tro → page 151
- van Doorn, J., Aust, F., Haaf, J. M., Stefan, A. M., and Wagenmakers, E.-J. (2021). Bayes factors for mixed models. *Computational Brain & Behavior* doi:10.1007/s42113-021-00113-2 → page 75
- Vaughn, C., Baese-Berk, M., and Idemaru, K. (2019). Re-examining phonetic variability in native and non-native speech. *Phonetica* 76, 327–358. doi:10.1159/000487269 → page 149
- Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* 27, 1413–1432. doi:10.1007/s11222-016-9709-3 → page 35
- Wade, T., Jongman, A., and Sereno, J. (2007). Effects of acoustic variability in the perceptual learning of non-native-accented speech sounds. *Phonetica* 64, 122–144. doi:10.1159/000107913 → page 148
- Whalen, D. H., Abramson, A. S., Lisker, L., and Mody, M. (1990). Gradient effects of fundamental frequency on stop consonant voicing judgments. *Phonetica* 47, 36–49. doi:10.1121/1.406678 → page 57

- Whalen, D. H., Abramson, A. S., Lisker, L., and Mody, M. (1993). *F0* gives voicing information even with unambiguous voice onset times. *The Journal of the Acoustical Society of America* 4, 2152–2159. doi:10.1121/1.406678 → pages 2, 57
- Whalen, D. H. and Levitt, A. G. (1995). The universality of intrinsic  $F_0$  of vowels. *Journal of Phonetics* 23, 349–366. doi:10.1016/S0095-4470(95)80165-0 → pages 36, 75, 129
- Winn, M. B. (2020). Manipulation of voice onset time in speech stimuli: A tutorial and flexible Praat script. *The Journal of the Acoustical Society of America* 147, 852–866. doi:10.1121/10.0000692 → pages 88, 90, 91
- Woods, K. J. P., Siegel, M. H., Traer, J., and McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics* 79, 2064–2072. doi:10.3758/s13414-017-1361-2 → page 96
- Wulff, D. U., Haslbeck, J. M. B., Kieslich, P. J., Henninger, F., and Schulte-Mecklenbeck, M. (2019). Mouse-tracking: Detecting types in movement trajectories. In *A handbook of process tracing methods*, eds. M. Schulte-Mecklenbeck, A. Kühberger, and J. G. Johnson (New York, NY: Routledge), chap. 9. 2 edn., 131–145. doi:10.4324/9781315160559 → page 5
- Xie, X. and Jaeger, T. F. (2020). Comparing non-native and native speech: Are L2 productions more variable? *The Journal of the Acoustical Society of America* 147, 3322–3347. doi:10.1121/10.0001141 → page 149
- Xu, C. X. and Xu, Y. (2003). Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Association* 33, 165–181. doi:10.1017/S0025100303001270 → pages 3, 12, 13, 14, 15, 16, 17, 18, 25, 56, 57, 58, 113
- Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics* 25, 61–83. doi:10.1006/jpho.1996.0034 → pages 13, 17, 49
- Yang, J., Zhang, Y., Li, A., and Xu, L. (2017). On the duration of Mandarin tones. In *Proceedings of Interspeech 2017*. 1407–1411. doi:10.21437/Interspeech.2017-29 → page 90
- Yu, A. C. and Zellou, G. (2019). Individual differences in language processing: Phonology. *Annual Review of Linguistics* 5, 131–150. doi:10.1146/annurev-linguistics-011516-033815 → page 202

- Yuan, J., Ryant, N., and Liberman, M. (2014). Automatic phonetic segmentation in Mandarin Chinese: Boundary models, glottal features and tone. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2539–2543 → pages 23, 26, 45
- Zee, E. (1980). The effect of aspiration on the  $f_0$  of the following vowel in Cantonese. *UCLA Working Papers in Phonetics* 49, 90–97 → page 51
- Zellers, M. and Post, B. (2009). Fundamental frequency and other prosodic cues to topic structure. In *Proceedings of Interfaces Discourse & Prosody 2009*, eds. H.-Y. Yoo and E. Delais-Roussarie. 377–386 → page 49
- Zellou, G. (2017). Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics* 61, 13–29. doi:10.1016/j.wocn.2016.12.002 → pages 59, 202
- Zhang, J. (2018). A comparison of tone normalization methods for language variation research. In *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation*. 823–831 → page 23

## Appendix A

# Language Background Questionnaire

### A.1 Language

1. Do you **currently** have a diagnosed speech disorder or hearing disability?
  - No
  - Yes, please specify: \_\_\_\_\_
2. Have you **previously** been treated for a speech disorder or hearing disability?
  - No
  - Yes, please specify: \_\_\_\_\_
3. What was the **first language** (or languages) you heard or you remember being exposed to from birth? Be as specific as you can with the name of the languages (e.g., Punjabi, Cantonese, Mandarin, Tagalog, Filipino). If you were exposed to more than one language from birth, please list them all.

Some of us heard or are exposed to a language that we no longer feel as if we really know. Please include these languages as well.

4. Do you still speak the language(s) you were exposed to since birth?

- If no, when do you think you stopped using it?

5. Please indicate all the languages **you have or have had experience with**. Again, be as specific as possible with the name of the language (e.g., Punjabi, Cantonese, Mandarin, Tagalog, Filipino). Please include languages you speak **natively** and ones you have **learned in other settings** (e.g., educational, self-study, etc.).
6. Regardless of your English proficiency, what variety or English do you think you speak? You can specify more than one. Please specify a dialect if you would like (e.g., Newfoundland English, Southern US English, etc.).

- American English
- Australian English
- British English
- Canadian English
- Filipino English
- Indian English
- Irish English
- Hong Kong English
- Jamaican English
- New Zealand English
- Scottish English
- Singaporean English
- South African English
- Other, please specify: \_\_\_\_\_

7. When were you first exposed to the language(s)? (If you started learning a language **from birth**, be sure to **put 0**.)

How would you rate your proficiency in reading, writing, speaking, and understanding?

We are following the Common European Framework for our language proficiency reference levels. The categories are defined as follows:

<b>No proficiency</b>	You have <b>no communicative ability</b> . Knowledge may be limited to memorized isolated words.
<b>Elementary</b>	You can understand and use <b>everyday expressions</b> and <b>basic phrases</b> . You can communicate simple, routine and immediate needs. For example, you can introduce yourself, order food, and ask for/give directions.
<b>Fair</b>	You can understand the <b>main points of clear input on familiar matters</b> . You can deal with topics that are likely to arise in social situations. For example, you can describe experiences, events, and your hopes and dreams.
<b>Good</b>	You can understand the <b>main ideas of both concrete and abstract topics</b> . For example, you can discuss social, economic and political matters. You are functional in a <b>work environment</b> . It is easy for your to interact with a native speaker spontaneously.
<b>Excellent</b>	You can express and understand <b>everything or almost everything</b> . For example, you can understand the language used in magazines, movies, TV, and radio programs targeted for an educated audience. You can follow a conversation between two native speakers with ease.

Please tick the appropriate option based on your self-assessed skill.

- Age
- Reading
- Writing
- Speaking
- Understanding

8. Which language do you **speak** most frequently? Draw the language from the left to the box on the right. If you speak multiple languages with equal frequency, draw them all.

9. Which language do you **hear** most frequently? Draw the language from the left to the box on the right. If you hear multiple languages with equal frequency, draw them all.
10. For each of your languages, how often do you currently speak/hear it and with whom? Please select the closest option (every day, several times a week, once/twice a week, once a month, once a year, never, N/A).

Please make sure to differentiate between listening and speaking.

- Speaking

- |                                  |                          |
|----------------------------------|--------------------------|
| - With parents: _____            | - With co-workers: _____ |
| - With grandparents: _____       | - At school: _____       |
| - With friends: _____            | - At home: _____         |
| - With significant others: _____ |                          |

- Listening

- |                                  |                          |
|----------------------------------|--------------------------|
| - With parents: _____            | - With co-workers: _____ |
| - With grandparents: _____       | - At school: _____       |
| - With friends: _____            | - At home: _____         |
| - With significant others: _____ |                          |

## A.2 You beyond Language

1. What is your age? \_\_\_\_\_
2. What year were you born? \_\_\_\_\_
3. What is your gender identification? \_\_\_\_\_
4. Are you right-handed or left-handed?
  - Right-handed
  - Left-handed

- Both

5. What is your racial or ethnic heritage? Check all that apply.

- First Nations
- Asian
- Pacific Islander
- Black
- White
- Hispanic
- South Asian
- Other, please specify: \_\_\_\_\_

6. What is the highest educational degree you have earned (or are in the process of learning)?

- High school
- Undergraduate degree
- Graduate degree
- Other, please specify: \_\_\_\_\_

7. Where have you lived? What age were you when you lived in each place?

Beginning with the place where you were born, please list each town or city (and country, if appropriate) you have lived in for **6 months or more** and your approximate age at each place.

Example:

0-10, Toronto, ON

11-20, Vancouver, BC

### **A.3 Your Caretakers**

1. Where were your caretakers born and raised?
2. What are your caretakers' first languages? Please specify for each caretaker.

## Appendix B

# Bilingual Language Profile Survey

### B.1 Biographical Information

This survey is called the Bilingual Language Profile, and it was created with support from the Center for Open Educational Resources and Language Learning at the University of Texas at Austin to better understand the profiles of bilingual speakers in diverse settings with diverse backgrounds. We would like to ask you to help us by answering the following questions concerning your language history, use, attitudes, and proficiency.

The survey consists of 19 questions and will take less than 10 minutes to complete. This is not a test, so there are no right or wrong answers. Please answer every question and give your answers sincerely.

1. Age: \_\_\_\_\_
2. What is your current place of residence (city/province/country)? \_\_\_\_\_
3. What is the highest educational degree you have earned or are in the process of earning?
  - High school

- Undergraduate degree
- Graduate degree
- Other, please specify: \_\_\_\_\_

## B.2 Language History

In this section, we would like you to answer some factual questions about your language history by placing a check in the appropriate box.

- At what age did you **start learning** the following languages?
  - Mandarin: [since birth], [1], [2], ..., [20+]
  - English: [since birth], [1], [2], ..., [20+]
- At what age did you **start to feel comfortable** using the following languages?
  - Mandarin: [as early as I can remember], [not yet], [1], [2], ..., [20+]
  - English: [as early as I can remember], [not yet], [1], [2], ..., [20+]
- How many years of **classes (grammar, history, math, etc.)** have you had in the following languages (primary school through university)?
  - Mandarin: [0], [1], [2], ..., [20+]
  - English: [0], [1], [2], ..., [20+]
- How many years have you spent in a **country/region** where the following languages are spoken?
  - Mandarin: [0], [1], [2], ..., [20+]
  - English: [0], [1], [2], ..., [20+]
- How many years have you spent in a **family** where the following languages are spoken?
  - Mandarin: [0], [1], [2], ..., [20+]

- English: [0], [1], [2], ..., [20+]

6. How many years have you spent in a **work environment** where the following languages are spoken?

- Mandarin: [0], [1], [2], ..., [20+]
- English: [0], [1], [2], ..., [20+]

### B.3 Language Use

In this section, we would like you to answer some questions about your language use by filling the percentage (0-100) in the appropriate box.

**Total use for all languages in a given question should equal 100%.**

1. In an average week, what percentage of the time do you use the following languages **with friends**?

- Mandarin: [0%], [10%], [20%], ..., [100%]
- English: [0%], [10%], [20%], ..., [100%]
- Other languages: [0%], [10%], [20%], ..., [100%]

2. In an average week, what percentage of the time do you use the following languages **with family**?

- Mandarin: [0%], [10%], [20%], ..., [100%]
- English: [0%], [10%], [20%], ..., [100%]
- Other languages: [0%], [10%], [20%], ..., [100%]

3. In an average week, what percentage of the time do you use the following languages **at school/work**?

- Mandarin: [0%], [10%], [20%], ..., [100%]
- English: [0%], [10%], [20%], ..., [100%]
- Other languages: [0%], [10%], [20%], ..., [100%]

4. When you **talk to yourself**, how often do you talk to yourself in the following languages?
  - Mandarin: [0%], [10%], [20%], ..., [100%]
  - English: [0%], [10%], [20%], ..., [100%]
  - Other languages: [0%], [10%], [20%], ..., [100%]
5. When you **count**, how often do you count in the following languages?
  - Mandarin: [0%], [10%], [20%], ..., [100%]
  - English: [0%], [10%], [20%], ..., [100%]
  - Other languages: [0%], [10%], [20%], ..., [100%]

## B.4 Language Proficiency

In this section, we would like you to rate your language proficiency by giving marks from 0 to 6.

1. How well do you **speak** the following languages?
  - Mandarin: [0 (not well at all)], [1], [2], [3], [4], [5], [6 (very well)]
  - English: [0 (not well at all)], [1], [2], [3], [4], [5], [6 (very well)]
2. How well do you **understand** the following languages?
  - Mandarin: [0 (not well at all)], [1], [2], [3], [4], [5], [6 (very well)]
  - English: [0 (not well at all)], [1], [2], [3], [4], [5], [6 (very well)]
3. How well do you **read** the following languages?
  - Mandarin: [0 (not well at all)], [1], [2], [3], [4], [5], [6 (very well)]
  - English: [0 (not well at all)], [1], [2], [3], [4], [5], [6 (very well)]
4. How well do you **write** the following languages?
  - Mandarin: [0 (not well at all)], [1], [2], [3], [4], [5], [6 (very well)]
  - English: [0 (not well at all)], [1], [2], [3], [4], [5], [6 (very well)]

## **B.5 Language Attitudes**

In this section, we would like you to respond to statements about language attitudes by giving marks from 0 to 6.

1. I feel like myself when I speak English/Mandarin?
  - Mandarin: [0 (totally disagree)], [1], [2], [3], [4], [5], [6 (totally agree)]
  - English: [0 (totally disagree)], [1], [2], [3], [4], [5], [6 (totally agree)]
2. I identify with an English/Mandarin-speaking culture.
  - Mandarin: [0 (totally disagree)], [1], [2], [3], [4], [5], [6 (totally agree)]
  - English: [0 (totally disagree)], [1], [2], [3], [4], [5], [6 (totally agree)]
3. It is important to me to use (or eventually use) English/Mandarin like a native speaker.
  - Mandarin: [0 (totally disagree)], [1], [2], [3], [4], [5], [6 (totally agree)]
  - English: [0 (totally disagree)], [1], [2], [3], [4], [5], [6 (totally agree)]
4. I want others to think I am a native speaker of English/Mandarin.
  - Mandarin: [0 (totally disagree)], [1], [2], [3], [4], [5], [6 (totally agree)]
  - English: [0 (totally disagree)], [1], [2], [3], [4], [5], [6 (totally agree)]

## Appendix C

# Supplementary Materials for Chapter 2

### C.1 Speaker Demographic Information

Table C.1: Speaker information.

ID	Gender	ID	Gender
CHJ	male	RUO	female
CHX	female	SHH	female
DIL	male	SUC	male
DOH	male	TIK	male
FAJ	female	WAJ	male
HAT	male	XIH	male
KOF	female	XIJ	male
LIS	male	XIN	female
MAK	male	XIY	male
OUT	male	XUL	female

## C.2 Summary Statistics of the Dataset

**Table C.2:** Summary statistics for consonant duration (ms) and F0 (scaled and standardized) by utterance position, lexical tone, vowel height, place of articulation (PoA), and voicing: mean, standard deviation, and number of tokens.

Position	Tone	Height	PoA	Voicing	c. dur. (ms)		F0 (scaled)		F0 (std.)		N
					Mean	SD	Mean	SD	Mean	SD	
initial	tone 1	high	labial	aspirated	90.00	20.00	1.01	.03	1.41	.95	5
				unaspirated	52.73	10.09	1.01	.06	.19	.73	11
			alveolar	aspirated	87.50	20.62	1.07	.12	.72	.94	4
				unaspirated	51.74	9.84	1.03	.05	1.17	.81	23
			velar	aspirated	75.00	7.07	1.09	.17	.86	1.25	2
				unaspirated	58.10	18.34	1.01	.08	.54	.66	22
	non-high	labial	aspirated	63.33	15.28	.96	.10	-.01	.31	3	
			unaspirated	49.39	11.19	1.06	.07	.73	.89	35	
		alveolar	aspirated	69.38	14.20	1.08	.08	.12	.74	85	
			unaspirated	48.41	11.67	1.04	.06	.71	.99	74	
		sonorant		–	–	1.01	.06	.15	.85	11	
			velar	aspirated	91.28	23.60	1.06	.07	.82	.81	11
	unaspirated		58.65	38.84	1.06	.07	.66	.87	55		
		tone 4	high	labial	aspirated	100.00	–	1.06	–	2.77	–
	unaspirated				52.88	11.39	1.04	.08	.64	1.05	139
	sonorant				–	–	1.01	.04	.73	.81	74
	alveolar			aspirated	70.00	42.43	1.13	.08	.86	2.53	2
				unaspirated	49.53	11.05	1.08	.08	.72	.88	124
sonorant				–	–	1.01	.04	.84	1.07	95	
velar	aspirated	85.00	21.21	.98	.11	.22	.58	2			
	unaspirated	60.66	14.48	1.03	.10	1.04	.96	9			
non-high	labial	aspirated	85.00	7.07	1.10	.03	-.44	.50	2		
		unaspirated	53.32	17.75	1.08	.06	.39	.90	52		
		sonorant	55.93	15.77	1.02	.05	.64	.94	28		

**Table C.2 continued:** Summary statistics for consonant duration (ms) and F0 (scaled and standardized) by utterance position, lexical tone, vowel height, place of articulation (PoA), and voicing: mean, standard deviation, and number of tokens.

Position	Tone	Height	PoA	Voicing	c. dur. (ms)		F0 (scaled)		F0 (std.)		N
					Mean	SD	Mean	SD	Mean	SD	
			alveolar	aspirated	102.00	10.95	1.10	.07	.88	.73	5
				unaspirated	49.46	10.17	1.09	.07	.46	.93	177
				sonorant	–	–	1.02	.04	.30	1.00	43
			velar	aspirated	85.37	12.98	1.11	.07	.58	.91	15
				unaspirated	60.79	15.46	1.08	.08	1.15	.87	38
medial	tone 1	high	labial	aspirated	97.64	27.99	1.05	.08	.19	1.00	36
				unaspirated	58.38	18.10	1.01	.06	–.10	.92	186
			alveolar	aspirated	61.64	27.13	1.02	.07	.14	.87	383
				unaspirated	55.99	14.52	1.00	.06	–.13	.92	382
				sonorant	–	–	.93	.04	–.04	.47	2
			velar	aspirated	103.33	5.77	1.12	.03	.09	.57	3
				unaspirated	62.65	18.00	1.01	.07	–.12	.83	490
		non-high	labial	aspirated	73.08	26.26	1.03	.07	–.15	.66	13
				unaspirated	60.53	18.96	1.03	.06	.07	.83	290
			alveolar	aspirated	75.44	28.78	1.01	.06	–.18	.82	251
				unaspirated	56.32	17.26	1.03	.07	.13	.99	331
				sonorant	–	–	1.00	.04	–.01	.89	143
			velar	aspirated	87.19	29.07	1.03	.07	–.11	.87	125
				unaspirated	62.21	18.67	1.01	.07	.08	.91	558
	tone 4	high	labial	aspirated	77.31	30.59	1.06	.08	–.04	.87	66
				unaspirated	60.89	19.77	1.05	.08	.11	1.00	813
				sonorant	–	–	1.04	.06	–.34	.95	476
			alveolar	aspirated	85.00	24.49	1.08	.06	–.33	.73	8
				unaspirated	51.55	13.08	1.07	.07	.09	1.04	1449
				sonorant	–	–	1.02	.06	–.16	1.10	1447
			velar	aspirated	69.07	26.67	1.08	.10	.02	.89	94

**Table C.2 continued:** Summary statistics for consonant duration (ms) and F0 (scaled and standardized) by utterance position, lexical tone, vowel height, place of articulation (PoA), and voicing: mean, standard deviation, and number of tokens.

Position	Tone	Height	PoA	Voicing	c. dur. (ms)		F0 (scaled)		F0 (std.)		N
					Mean	SD	Mean	SD	Mean	SD	
		non-high	labial	unaspirated	54.22	17.33	1.06	.07	.24	.94	218
				aspirated	84.69	28.09	1.07	.09	-.12	.74	88
		non-high	labial	unaspirated	55.50	16.99	1.08	.07	-.19	.96	567
				sonorant	-	-	1.03	.05	.20	.97	212
			alveolar	aspirated	81.10	31.27	1.07	.09	.02	.81	139
				unaspirated	50.43	13.33	1.08	.06	-.20	1.02	1429
		non-high	alveolar	sonorant	-	-	1.04	.06	-.06	.96	344
				velar	aspirated	88.03	28.95	1.11	.09	-.03	.89
			unaspirated	51.30	15.79	1.05	.08	-.25	.95	692	

### C.3 Output of Models with Weighted Effect Coding

**Table C.3:** Marginal posterior summaries for population-level parameters from a scaled-F0 model with weighted effect coding. The model was fit with the following structure: F0 *sim* position + height + tone + voicing + PoA + (1 + position + height + tone + voicing + PoA || speaker) + (1 + position || word). The variables **position** (UTTERANCE-INITIAL = 1, UTTERANCE-MEDIAL =  $-1.10$ ), **height** (HIGH = 1, NON-HIGH =  $-1.10$ ), **tone** (TONE 1 = 1, TONE 4 =  $-0.39$ ), and **PoA** (LABIAL =  $[1, 0]$ , ALVEOLAR =  $[0, 1]$ , VELAR =  $[-1.25, -2.80]$ ) were weighted effect coded.

Parameter	Mean	SD	89% CrI	$p(\text{dir.})$
intercept	1.0452	.0035	[1.0393, 1.0506]	$p(\beta > 0) = 1.00$
utt-init	.0028	.0051	$[-.0054, .0107]$	$p(\beta > 0) = .72$
high	$-.0035$	.0020	$[-.0068, -.0002]$	$p(\beta < 0) = .96$
tone 1	$-.0312$	.0036	$[-.0368, -.0252]$	$p(\beta < 0) = 1.00$
asp – unasp	.0107	.0071	$[-.0005, .0222]$	$p(\beta > 0) = .94$
unasp – son	.0364	.0064	$ [.0263, .0467]$	$p(\beta > 0) = 1.00$
labial	.0016	.0031	$[-.0035, .0065]$	$p(\beta > 0) = .70$
alveolar	$-.0013$	.0022	$[-.0047, .0023]$	$p(\beta < 0) = .73$

**Table C.4:** Marginal posterior summaries for population-level parameters from a standardized-F0 model with weighted effect coding. The model was fit with the following structure: F0 *sim* position + height + tone + voicing + PoA + (1 + position + height + tone + voicing + PoA || speaker) + (1 + position || word). The variables **position** (UTTERANCE-INITIAL = 1, UTTERANCE-MEDIAL = -1.10), **height** (HIGH = 1, NON-HIGH = -1.10), **tone** (TONE 1 = 1, TONE 4 = -0.39), and **PoA** (LABIAL = [1,0], ALVEOLAR = [0,1], VELAR = [-1.25, -2.80]) were weighted effect coded.

Parameter	Mean	SD	89% CrI	$p(\text{dir.})$
intercept	.01	.04	[-.04, .07]	$p(\beta > 0) = .65$
utt-init	.63	.06	[.53, .73]	$p(\beta > 0) = 1.00$
high	.04	.03	[-.01, .09]	$p(\beta > 0) = .89$
tone 1	-.01	.05	[-.10, .08]	$p(\beta < 0) = .57$
asp – unasp	-.10	.08	[-.23, .03]	$p(\beta < 0) = .89$
unasp – son	.19	.10	[.03, .35]	$p(\beta > 0) = .97$
labial	.00	.06	[-.09, .09]	$p(\beta > 0) = .50$
alveolar	-.02	.03	[-.07, .03]	$p(\beta < 0) = .76$

## C.4 Posterior Summaries of Individual-Level Parameters

**Table C.5:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level. ① = intercept, ② =  $\text{utt-init} - (\text{utt-init} + \text{utt-med})/2$ , ③ =  $\text{high} - (\text{high} + \text{non-high})/2$ , ④ =  $\text{tone 1} - (\text{tone 1} + \text{tone 4})/2$ , ⑤ =  $\text{asp} - \text{unasp}$ , ⑥ =  $\text{unasp} - \text{son}$ .

Speaker	①	②	③	④	⑤	⑥
CHJ	1.0325 (.0060) [1.0229, 1.0422]	−.0027 (.0046) [−.0101, .0046]	−.0021 (.0035) [−.0075, .0035]	−.0255 (.0040) [−.0318, −.0192]	.0242 (.0116) [.0061, .0431]	.0264 (.0080) [.0136, .0390]
CHX	1.0397 (.0048) [1.0319, 1.0475]	.0062 (.0038) [.0002, .0122]	−.0010 (.0030) [−.0057, .0038]	−.0132 (.0031) [−.0182, −.0080]	.0133 (.0080) [.0006, .0259]	.0358 (.0076) [.0237, .0476]
DIL	1.0586 (.0057) [1.0495, 1.0676]	.0134 (.0043) [.0067, .0203]	−.0021 (.0035) [−.0076, .0035]	−.0112 (.0039) [−.0174, −.0048]	.0375 (.0106) [.0205, .0544]	.0540 (.0090) [.0395, .0684]
DOH	1.0258 (.0038) [1.0198, 1.0318]	.0088 (.0028) [.0043, .0133]	−.0121 (.0030) [−.0168, −.0072]	−.0164 (.0026) [−.0206, −.0122]	−.0267 (.0075) [−.0384, −.0147]	.0348 (.0060) [.0253, .0443]
FAJ	1.0227 (.0061) [1.0129, 1.0324]	−.0018 (.0054) [−.0104, .0068]	.0017 (.0030) [−.0032, .0066]	−.0240 (.0031) [−.0289, −.0192]	.0075 (.0083) [−.0058, .0207]	.0506 (.0070) [.0395, .0615]
HAT	1.0456 (.0058) [1.0366, 1.0549]	−.0007 (.0048) [−.0083, .0068]	−.0054 (.0032) [−.0106, −.0005]	−.0223 (.0035) [−.0277, −.0166]	.0185 (.0094) [.0039, .0332]	.0476 (.0077) [.0354, .0597]
KOF	1.0373 (.0063) [1.0270, 1.0476]	.0026 (.0047) [−.0048, .0103]	−.0035 (.0037) [−.0093, .0022]	−.0150 (.0042) [−.0216, −.0083]	.0218 (.0122) [.0019, .0414]	.0313 (.0100) [.0156, .0474]
LIS	1.0658 (.0054) [1.0570, 1.0745]	.0093 (.0040) [.0028, .0159]	−.0028 (.0034) [−.0081, .0026]	−.0203 (.0036) [−.0261, −.0146]	−.0048 (.0097) [−.0201, .0105]	.0868 (.0098) [.0712, .1025]

**Table C.5 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level. ① = intercept, ② = utt-init - (utt-init + utt-med)/2, ③ = high - (high + non-high)/2, ④ = tone 1 - (tone 1 + tone 4)/2, ⑤ = asp - unasp, ⑥ = unasp - son.

Speaker	①	②	③	④	⑤	⑥
MAK	1.0436 (.0064) [1.0333, 1.0540]	.0065 (.0052) [-.0018, .0149]	-.0042 (.0037) [-.0102, .0016]	-.0211 (.0048) [-.0288, -.0133]	-.0219 (.0127) [-.0425, -.0018]	.0620 (.0102) [.0459, .0783]
OUT	1.0285 (.0052) [1.0203, 1.0370]	.0030 (.0041) [-.0035, .0095]	-.0087 (.0038) [-.0146, -.0027]	-.0241 (.0037) [-.0301, -.0182]	-.0240 (.0105) [-.0411, -.0071]	.0203 (.0096) [.0050, .0355]
RUO	1.0133 (.0071) [1.0020, 1.0244]	-.0030 (.0048) [-.0107, .0045]	-.0038 (.0037) [-.0097, .0021]	-.0215 (.0042) [-.0284, -.0148]	.0157 (.0138) [-.0057, .0377]	.0178 (.0085) [.0042, .0313]
SHH	1.0410 (.0049) [1.0330, 1.0488]	-.0027 (.0039) [-.0089, .0033]	.0020 (.0033) [-.0031, .0074]	-.0199 (.0032) [-.0250, -.0148]	.0280 (.0094) [.0130, .0430]	.0402 (.0072) [.0284, .0515]
SUC	1.0490 (.0057) [1.0402, 1.0582]	.0147 (.0049) [.0069, .0227]	-.0075 (.0034) [-.0130, -.0022]	-.0138 (.0036) [-.0195, -.0080]	.0074 (.0089) [-.0065, .0216]	.0401 (.0083) [.0266, .0535]
TIK	1.0297 (.0084) [1.0159, 1.0431]	-.0058 (.0062) [-.0157, .0036]	-.0026 (.0045) [-.0097, .0047]	-.0220 (.0058) [-.0316, -.0127]	.0155 (.0179) [-.0132, .0443]	.0363 (.0126) [.0164, .0556]
WAJ	1.0312 (.0047) [1.0237, 1.0386]	-.0063 (.0041) [-.0127, .0000]	-.0004 (.0026) [-.0045, .0039]	-.0328 (.0026) [-.0368, -.0287]	.0439 (.0065) [.0334, .0542]	.0340 (.0063) [.0241, .0441]
XIH	1.0314 (.0045) [1.0244, 1.0386]	.0000 (.0037) [-.0058, .0059]	-.0026 (.0028) [-.0070, .0019]	-.0286 (.0028) [-.0331, -.0241]	.0358 (.0068) [.0251, .0466]	.0332 (.0074) [.0216, .0449]
XIJ	1.0418 (.0062)	.0065 (.0051)	-.0037 (.0034)	-.0176 (.0039)	-.0258 (.0113)	.0468 (.0082)

**Table C.5 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level. ① = intercept, ② =  $\text{utt-init} - (\text{utt-init} + \text{utt-med})/2$ , ③ =  $\text{high} - (\text{high} + \text{non-high})/2$ , ④ =  $\text{tone 1} - (\text{tone 1} + \text{tone 4})/2$ , ⑤ =  $\text{asp} - \text{unasp}$ , ⑥ =  $\text{unasp} - \text{son}$ .

Speaker	①	②	③	④	⑤	⑥
	[1.0320, 1.0517]	[−.0016, .0145]	[−.0092, .0017]	[−.0238, −.0114]	[−.0438, −.0075]	[.0339, .0602]
XIN	1.0549 (.0064)	.0002 (.0049)	−.0020 (.0036)	−.0192 (.0038)	.0239 (.0123)	.0490 (.0083)
	[1.0449, 1.0652]	[−.0076, .0080]	[−.0077, .0039]	[−.0253, −.0132]	[.0044, .0436]	[.0356, .0623]
XIY	1.0558 (.0049)	−.0025 (.0040)	.0017 (.0032)	−.0257 (.0030)	.0461 (.0086)	.0353 (.0075)
	[1.0480, 1.0636]	[−.0089, .0037]	[−.0033, .0068]	[−.0304, −.0208]	[.0323, .0599]	[.0236, .0473]
XUL	1.0126 (.0056)	−.0062 (.0047)	−.0025 (.0030)	−.0281 (.0031)	.0090 (.0083)	.0301 (.0073)
	[1.0034, 1.0215]	[−.0139, .0012]	[−.0075, .0022]	[−.0331, −.0231]	[−.0042, .0222]	[.0182, .0415]

244

**Table C.6:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level. ⑦ =  $\text{labial} - (\text{labial} + \text{alveolar} + \text{velar})/3$ , ⑧ =  $\text{alveolar} - (\text{labial} + \text{alveolar} + \text{velar})/3$ , ⑨ =  $[\text{asp} - \text{unasp}] \times [\text{high} - (\text{high} + \text{non-high})/2]$ , ⑩ =  $[\text{unasp} - \text{son}] \times [\text{high} - (\text{high} + \text{non-high})/2]$ .

Speaker	⑦	⑧	⑨	⑩
CHJ	.0015 (.0037)	−.0026 (.0046)	.0207 (.0093)	−.0066 (.0067)
	[−.0042, .0077]	[−.0100, .0046]	[.0066, .0358]	[−.0170, .0035]
CHX	−.0011 (.0034)	−.0051 (.0037)	.0092 (.0068)	.0015 (.0064)
	[−.0068, .0041]	[−.0110, .0008]	[−.0017, .0199]	[−.0087, .0120]

**Table C.6 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level. ⑦ = labial – (labial + alveolar + velar)/3, ⑧ = alveolar – (labial + alveolar + velar)/3, ⑨ = [asp – unasp] × [high – (high + non-high)/2], ⑩ = [unasp – son] × [high – (high + non-high)/2].

Speaker	⑦	⑧	⑨	⑩
DIL	.0000 (.0036) [–.0058, .0058]	–.0023 (.0044) [–.0095, .0048]	.0148 (.0081) [.0021, .0280]	–.0021 (.0071) [–.0133, .0094]
DOH	.0004 (.0031) [–.0047, .0052]	.0285 (.0033) [.0232, .0337]	.0205 (.0071) [.0094, .0316]	–.0152 (.0056) [–.0242, –.0063]
FAJ	.0020 (.0035) [–.0033, .0078]	–.0039 (.0038) [–.0102, .0023]	.0225 (.0076) [.0106, .0349]	–.0162 (.0064) [–.0264, –.0058]
HAT	–.0015 (.0036) [–.0076, .0039]	–.0023 (.0042) [–.0089, .0044]	.0094 (.0074) [–.0023, .0214]	–.0048 (.0065) [–.0149, .0057]
KOF	–.0018 (.0039) [–.0083, .0041]	–.0043 (.0051) [–.0123, .0038]	.0121 (.0088) [–.0017, .0262]	–.0035 (.0075) [–.0153, .0083]
LIS	–.0008 (.0037) [–.0070, .0050]	–.0103 (.0044) [–.0171, –.0033]	.0098 (.0077) [–.0023, .0219]	–.0064 (.0072) [–.0180, .0046]
MAK	.0010 (.0039) [–.0050, .0075]	.0031 (.0054) [–.0054, .0119]	.0043 (.0086) [–.0099, .0176]	–.0069 (.0075) [–.0189, .0051]
OUT	–.0007 (.0037) [–.0069, .0051]	–.0028 (.0046) [–.0100, .0047]	.0055 (.0079) [–.0076, .0177]	–.0095 (.0075) [–.0219, .0022]
RUO	–.0004 (.0037)	–.0081 (.0051)	.0113 (.0090)	–.0036 (.0069)

**Table C.6 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level. ⑦ = labial – (labial + alveolar + velar)/3, ⑧ = alveolar – (labial + alveolar + velar)/3, ⑨ = [asp – unasp] × [high – (high + non-high)/2], ⑩ = [unasp – son] × [high – (high + non-high)/2].

Speaker	⑦	⑧	⑨	⑩
SHH	[–.0066, .0054] .0034 (.0039)	[–.0162, –.0001] –.0012 (.0038)	[–.0028, .0258] .0033 (.0077)	[–.0144, .0077] .0080 (.0067)
SUC	[–.0025, .0098] .0008 (.0035)	[–.0076, .0049] –.0064 (.0041)	[–.0093, .0153] –.0037 (.0082)	[–.0026, .0188] –.0076 (.0068)
TIK	[–.0047, .0066] –.0002 (.0040)	[–.0130, .0002] .0138 (.0072)	[–.0173, .0094] .0195 (.0108)	[–.0183, .0032] –.0065 (.0088)
WAJ	[–.0067, .0060] –.0009 (.0031)	[.0025, .0255] –.0030 (.0030)	[.0035, .0379] .0044 (.0058)	[–.0207, .0073] –.0059 (.0055)
XIH	[–.0060, .0040] .0027 (.0035)	[–.0078, .0018] .0042 (.0034)	[–.0049, .0135] .0041 (.0062)	[–.0146, .0028] .0002 (.0061)
XIJ	[–.0027, .0086] .0005 (.0035)	[–.0012, .0095] –.0158 (.0045)	[–.0059, .0138] .0106 (.0079)	[–.0096, .0099] –.0128 (.0073)
XIN	[–.0051, .0059] –.0026 (.0040)	[–.0229, –.0087] .0002 (.0046)	[–.0017, .0235] .0042 (.0092)	[–.0246, –.0015] –.0036 (.0067)
XIY	[–.0094, .0032] –.0024 (.0036)	[–.0072, .0074] .0029 (.0039)	[–.0110, .0185] .0120 (.0071)	[–.0142, .0072] .0022 (.0065)
	[–.0081, .0031]	[–.0033, .0090]	[.0009, .0231]	[–.0077, .0130]

**Table C.6 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level. ⑦ = labial – (labial + alveolar + velar)/3, ⑧ = alveolar – (labial + alveolar + velar)/3, ⑨ = [asp – unasp] × [high – (high + non-high)/2], ⑩ = [unasp – son] × [high – (high + non-high)/2].

Speaker	⑦	⑧	⑨	⑩
XUL	.0003 (.0033) [–.0050, .0057]	.0088 (.0036) [.0032, .0146]	.0085 (.0069) [–.0026, .0193]	–.0069 (.0062) [–.0170, .0030]

**Table C.7:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level. ① = intercept, ② = utt-init – (utt-init + utt-med)/2, ③ = high – (high + non-high)/2, ④ = tone 1 – (tone 1 + tone 4)/2, ⑤ = asp – unasp, ⑥ = unasp – son, ⑦ = labial – (labial + alveolar + velar)/3, ⑧ = alveolar – (labial + alveolar + velar)/3.

Speaker	①	②	③	④	⑤	⑥	⑦	⑧
CHJ	.30 (.05) [.23, .38]	.35 (.04) [.30, .41]	.05 (.04) [–.02, .12]	–.03 (.04) [–.10, .04]	–.10 (.11) [–.28, .09]	.28 (.12) [.10, .47]	.03 (.07) [–.09, .14]	–.04 (.05) [–.12, .04]
CHX	.31 (.05) [.24, .38]	.34 (.04) [.28, .40]	.04 (.04) [–.02, .10]	–.03 (.04) [–.10, .04]	–.03 (.10) [–.19, .13]	.26 (.11) [.08, .43]	.00 (.07) [–.10, .11]	–.05 (.05) [–.13, .03]
DIL	.31 (.05) [.23, .39]	.35 (.04) [.30, .41]	.04 (.05) [–.04, .11]	–.02 (.04) [–.09, .05]	–.04 (.11) [–.22, .15]	.34 (.12) [.15, .54]	.00 (.07) [–.11, .12]	–.03 (.05) [–.12, .05]
DOH	.29 (.05) [.21, .36]	.38 (.04) [.32, .44]	.02 (.04) [–.05, .08]	–.02 (.04) [–.08, .04]	–.24 (.11) [–.42, –.06]	.06 (.10) [–.10, .22]	.01 (.06) [–.08, .10]	–.02 (.05) [–.10, .06]

**Table C.7 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level. ① = intercept, ② = utt-init - (utt-init + utt-med)/2, ③ = high - (high + non-high)/2, ④ = tone 1 - (tone 1 + tone 4)/2, ⑤ = asp - unasp, ⑥ = unasp - son, ⑦ = labial - (labial + alveolar + velar)/3, ⑧ = alveolar - (labial + alveolar + velar)/3.

Speaker	①	②	③	④	⑤	⑥	⑦	⑧
FAJ	.32 (.05) [.25, .40]	.36 (.04) [.30, .41]	.08 (.04) [.01, .14]	.00 (.04) [-.06, .07]	-.13 (.10) [-.29, .04]	.31 (.11) [.14, .48]	.05 (.07) [-.05, .16]	-.03 (.05) [-.11, .06]
HAT	.31 (.05) [.24, .39]	.34 (.04) [.28, .40]	.04 (.04) [-.04, .11]	-.02 (.04) [-.09, .05]	-.09 (.11) [-.26, .08]	.22 (.11) [.05, .40]	.01 (.07) [-.10, .12]	-.04 (.05) [-.12, .04]
KOF	.31 (.05) [.23, .39]	.35 (.04) [.29, .41]	.02 (.05) [-.06, .10]	.00 (.04) [-.07, .07]	.01 (.13) [-.18, .23]	.15 (.13) [-.05, .34]	-.09 (.08) [-.22, .03]	-.03 (.05) [-.12, .05]
LIS	.31 (.05) [.23, .38]	.35 (.04) [.30, .41]	.08 (.05) [.01, .15]	-.02 (.04) [-.09, .04]	-.07 (.11) [-.23, .11]	.27 (.12) [.09, .46]	.06 (.07) [-.05, .18]	-.06 (.05) [-.15, .03]
MAK	.31 (.05) [.24, .39]	.35 (.04) [.29, .41]	.07 (.05) [-.01, .14]	-.03 (.05) [-.11, .04]	-.20 (.13) [-.41, .00]	.17 (.12) [-.03, .37]	-.07 (.08) [-.20, .05]	-.03 (.05) [-.11, .06]
OUT	.30 (.05) [.22, .37]	.35 (.04) [.29, .40]	.09 (.05) [.01, .16]	-.02 (.04) [-.09, .04]	-.18 (.12) [-.38, .00]	.17 (.12) [-.03, .36]	-.02 (.07) [-.14, .09]	-.04 (.05) [-.12, .04]
RUO	.31 (.05) [.23, .39]	.34 (.04) [.28, .40]	.11 (.05) [.03, .19]	-.02 (.04) [-.09, .05]	-.13 (.12) [-.32, .06]	.04 (.12) [-.15, .23]	-.10 (.08) [-.23, .03]	-.01 (.06) [-.10, .09]
SHH	.31 (.05) [.23, .38]	.34 (.04) [.27, .39]	.08 (.04) [.01, .15]	-.02 (.04) [-.09, .04]	-.07 (.11) [-.24, .10]	.27 (.11) [.10, .45]	.09 (.07) [-.02, .20]	-.05 (.05) [-.13, .03]

**Table C.7 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level. ① = intercept, ② = utt-init - (utt-init + utt-med)/2, ③ = high - (high + non-high)/2, ④ = tone 1 - (tone 1 + tone 4)/2, ⑤ = asp - unasp, ⑥ = unasp - son, ⑦ = labial - (labial + alveolar + velar)/3, ⑧ = alveolar - (labial + alveolar + velar)/3.

Speaker	①	②	③	④	⑤	⑥	⑦	⑧
SUC	.32 (.05) [.25, .40]	.36 (.04) [.30, .42]	.08 (.04) [.01, .15]	-.02 (.04) [-.09, .05]	-.09 (.11) [-.26, .08]	.32 (.12) [.14, .51]	.01 (.07) [-.10, .12]	-.05 (.05) [-.13, .03]
TIK	.31 (.05) [.23, .39]	.35 (.04) [.29, .42]	.04 (.05) [-.04, .12]	.00 (.05) [-.08, .07]	-.11 (.13) [-.31, .10]	.11 (.14) [-.12, .33]	-.01 (.09) [-.15, .13]	-.02 (.06) [-.11, .08]
WAJ	.33 (.05) [.26, .41]	.36 (.04) [.30, .41]	.03 (.04) [-.03, .09]	-.01 (.04) [-.07, .06]	-.04 (.09) [-.19, .11]	.12 (.10) [-.04, .28]	-.01 (.06) [-.10, .08]	-.06 (.05) [-.14, .02]
XIH	.30 (.05) [.22, .37]	.37 (.04) [.32, .43]	.04 (.04) [-.02, .10]	-.03 (.04) [-.09, .04]	-.10 (.10) [-.26, .05]	.13 (.10) [-.04, .30]	-.14 (.06) [-.25, -.04]	-.02 (.05) [-.10, .06]
XIJ	.33 (.05) [.25, .41]	.35 (.04) [.30, .41]	.12 (.05) [.04, .20]	-.02 (.04) [-.09, .04]	-.19 (.12) [-.38, -.01]	.15 (.12) [-.04, .34]	-.08 (.07) [-.20, .03]	-.03 (.05) [-.11, .05]
XIN	.31 (.05) [.24, .39]	.33 (.04) [.25, .39]	.10 (.05) [.03, .18]	-.02 (.04) [-.09, .05]	-.07 (.12) [-.26, .11]	.07 (.12) [-.12, .25]	.00 (.07) [-.12, .11]	-.07 (.06) [-.16, .02]
XIY	.32 (.05) [.24, .39]	.35 (.03) [.30, .41]	.04 (.04) [-.03, .10]	.03 (.05) [-.05, .11]	-.01 (.11) [-.18, .17]	.14 (.11) [-.03, .31]	-.02 (.06) [-.12, .09]	-.05 (.05) [-.13, .03]
XUL	.32 (.05) [.24, .40]	.37 (.04) [.31, .43]	.07 (.04) [.00, .13]	.00 (.04) [-.07, .06]	-.16 (.10) [-.33, .01]	.16 (.11) [-.01, .33]	-.08 (.06) [-.19, .02]	-.03 (.05) [-.11, .05]

## Appendix D

# Supplementary Materials for Chapter 3

### D.1 Participant Demographic Information

**Table D.1:** Demographic information of the L1 Mandarin-L2 English participants. BLP = Bilingual Language Profile Score (see Section 6.4.2); Am. = American English; Br. = British English; Ca. = Canadian English.

ID	Gender	Age	BLP	Grow-up place	Eng. variety	Other lang.
1	F	20	72.6	Hangzhou	Am./Ca.	
2	F	27	56.0	Shandong	Ca.	
3	F	23	88.3	Yueyang	Ca.	Cantonese
4	M	26	112.0	Beijing	Ca.	Russian
5	F	22	88.7	Guiyang	—	
6	M	21	93.6	Changsha	Ca.	
7	F	21	70.1	Chengdu	Am.	
8	M	22	62.6	Beijing	Am.	
9	F	21	59.8	Shanghai	Am.	Shanghainese
10	M	22	65.6	Suzhou	Am./Br./Ca.	
11	F	21	81.4	Nanjing	Am.	

**Table D.1 continued:** Demographic information of the L1 Mandarin-L2 English participants. BLP = Bilingual Language Profile Score (see Section 6.4.2); Am. = American English; Br. = British English; Ca. = Canadian English.

<b>ID</b>	<b>Gender</b>	<b>Age</b>	<b>BLP</b>	<b>Grow-up place</b>	<b>Eng. variety</b>	<b>Other lang.</b>
12	F	21	107.2	Wenzhou	Am./Ca.	
13	M	20	24.2	Beijing	Ca.	
14	F	20	56.5	Qingdao	Ca.	Japanese
15	F	20	92.5	Xiamen	Ca.	
16	F	20	70.2	Shanghai	Am./Br./Ca.	
17	F	20	77.1	Shanghai	Am./Ca.	
18	M	21	102.0	Nanjing	Ca.	
19	F	19	104.0	Baicheng	Ca.	
20	M	21	71.8	Shenzhen	Am./Ca.	
21	M	20	55.1	Ningbo	Ca.	Spanish
22	M	19	48.0	[China]	Am./Br.	
23	M	19	91.0	Beijing	Am./Ca.	
24	M	18	70.1	Lanzhou	Am.	
25	F	18	70.7	Beijing	Am.	

## D.2 Statistical Model Specification

This section offers the mathematical formulations for the statistical models covered in Section 3.3.7 and Section 3.4.6.

### D.2.1 Production Model for Post-Stop F0

Likelihood:

$$\begin{aligned}
 F0_i^j = & (\beta_0 + u_0^j) + \\
 & (\beta_1 + u_1^j) \cdot \text{HIGH}_i + && \text{(vowel height)} \\
 & (\beta_2 + u_2^j) \cdot [\text{ENG vs. MAN T1/T4}]_i + \\
 & (\beta_3 + u_3^j) \cdot [\text{MAN T1 vs. MAN T4}]_i + && \text{(language/tone)} \\
 & (\beta_4 + u_4^j) \cdot [\text{ASP vs. UNASP}]_i + \\
 & (\beta_5 + u_5^j) \cdot [\text{UNASP vs. SON}]_i + && \text{(voicing)} \\
 & (\beta_6 + u_6^j) \cdot \text{HIGH}_i \times [\text{ASP vs. UNASP}]_i + \\
 & (\beta_7 + u_7^j) \cdot \text{HIGH}_i \times [\text{UNASP vs. SON}]_i + && \text{(vowel height} \times \text{voicing)} \\
 & (\beta_8 + u_8^j) \cdot [\text{ENG vs. MAN T1/T4}]_i \times [\text{ASP vs. UNASP}]_i + \\
 & (\beta_9 + u_9^j) \cdot [\text{MAN T1 vs. MAN T4}]_i \times [\text{ASP vs. UNASP}]_i + \\
 & (\beta_{10} + u_{10}^j) \cdot [\text{ENG vs. MAN T1/T4}]_i \times [\text{UNASP vs. SON}]_i + \\
 & (\beta_{11} + u_{11}^j) \cdot [\text{MAN T1 vs. MAN T4}]_i \times [\text{UNASP vs. SON}]_i + && \text{(language/tone} \times \text{voicing)} \\
 & \varepsilon && \text{(error term)} \\
 \varepsilon \sim & \mathcal{N}(0, \sigma^2)
 \end{aligned}$$

Priors:

$$\beta_0, \beta_1, \dots, \beta_{11} \sim \mathcal{N}(0, 5)$$

$$\begin{bmatrix} u_0^j \\ u_1^j \\ \vdots \\ u_{11}^j \end{bmatrix} \sim \mathcal{N}(0, S), j = 1, 2, \dots, 25$$

$$S = \begin{bmatrix} \sigma_0 & 0 & \cdots & 0 \\ 0 & \sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \sigma_{11} \end{bmatrix} R \begin{bmatrix} \sigma_0 & 0 & \cdots & 0 \\ 0 & \sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \sigma_{11} \end{bmatrix}$$

$$R \sim \text{LKJCorr}(1)$$

$$\sigma \sim \text{Exponential}(1)$$

$$\sigma_0, \sigma_1, \dots, \sigma_{11} \sim \text{Exponential}(1)$$

## D.2.2 Perceptual Model for Post-Stop F0 Weight

Likelihood:

$$\begin{aligned}
 \ell=\text{Eng}/\text{p}/ \text{response}_i^j \sim \text{Bernoulli}(& \\
 & \ell=\text{Eng} P_{\text{guess}}^j \times 0.25 + \\
 & (1 - \ell=\text{Eng} P_{\text{guess}}^j) \times \text{logit}^{-1}(& \\
 & (\beta_0 + u_0^j) + & \\
 & (\beta_1 + u_1^j) \cdot \text{VOT}_i + & \text{(VOT)} \\
 & (\beta_2 + u_2^j) \cdot \text{F0}_i + & \text{(F0)} \\
 & (\beta_3 + u_3^j) \cdot \text{TONE } 1_i + & \text{(tone)} \\
 & )) & 
 \end{aligned}$$

$$\begin{aligned}
 \ell=\text{Man}/\text{p}/ \text{response}_i^j \sim \text{Bernoulli}(& \\
 & \ell=\text{Man} P_{\text{guess}}^j \times 0.25 + \\
 & (1 - \ell=\text{Man} P_{\text{guess}}^j) \times \text{logit}^{-1}(& \\
 & (\beta_4 + u_4^j) + & \\
 & (\beta_5 + u_5^j) \cdot \text{VOT}_i + & \text{(VOT)} \\
 & (\beta_6 + u_6^j) \cdot \text{F0}_i + & \text{(F0)} \\
 & (\beta_7 + u_7^j) \cdot \text{TONE } 1_i + & \text{(tone)} \\
 & )) & 
 \end{aligned}$$

Priors:

$${}^{\ell=\text{Eng}}P_{\text{guess}}^j, {}^{\ell=\text{Man}}P_{\text{guess}}^j \sim \text{Unif}(0, 1), j = 1, 2, \dots, 25$$

$$\beta_0, \beta_1, \dots, \beta_7 \sim \mathcal{N}(0, 10)$$

$$\begin{bmatrix} u_0^j \\ u_1^j \\ \vdots \\ u_7^j \end{bmatrix} \sim \mathcal{N}(0, S), j = 1, 2, \dots, 25$$

$$S = \begin{bmatrix} \sigma_0 & 0 & \cdots & 0 \\ 0 & \sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \sigma_7 \end{bmatrix} R \begin{bmatrix} \sigma_0 & 0 & \cdots & 0 \\ 0 & \sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \sigma_7 \end{bmatrix}$$

$$\sigma_0, \sigma_1, \dots, \sigma_7 \sim \text{Exponential}(1)$$

$$R \sim \text{LKJCorr}(1)$$

### D.3 Individual-Level Posterior Parameter Summaries

**Table D.2:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level. ① = intercept, ② = high – (high + low)/2, ③ = Eng – (tone1 + tone4)/2, ④ = tone1 – tone4, ⑤ = asp – unasp, ⑥ = unasp – son, ⑦ = [asp – unasp] × [high – (high + low)/2], ⑧ = [unasp – son] × [high – (high + low)/2].

Par.	①	②	③	④	⑤	⑥	⑦	⑧
MS01	–.20 (.05) [–.27, –.11]	.36 (.06) [.26, .45]	–.49 (.11) [–.67, –.30]	–1.35 (.15) [–1.60, –1.11]	.68 (.12) [.48, .87]	.36 (.14) [.14, .58]	.10 (.10) [–.04, .27]	–.03 (.10) [–.21, .12]
MS02	–.22 (.06) [–.31, –.13]	.32 (.07) [.21, .43]	–.41 (.17) [–.66, –.13]	–.63 (.15) [–.87, –.38]	.71 (.19) [.41, 1.01]	.56 (.27) [.16, .99]	–.08 (.13) [–.32, .09]	.15 (.14) [–.02, .40]
MS03	–.22 (.05) [–.30, –.14]	.39 (.07) [.28, .50]	–.41 (.16) [–.65, –.15]	–1.34 (.15) [–1.60, –1.10]	.66 (.18) [.38, .94]	.43 (.22) [.10, .79]	.08 (.10) [–.07, .24]	.06 (.10) [–.08, .24]
MS04	–.22 (.05) [–.31, –.14]	.34 (.07) [.23, .45]	–.42 (.16) [–.65, –.16]	–1.30 (.14) [–1.53, –1.07]	.67 (.18) [.39, .93]	.22 (.23) [–.15, .57]	.07 (.10) [–.08, .23]	.03 (.10) [–.12, .19]
MS05	–.22 (.06) [–.31, –.13]	.32 (.07) [.20, .43]	–.41 (.17) [–.65, –.13]	–.78 (.15) [–1.01, –.52]	.83 (.21) [.51, 1.19]	.41 (.23) [.07, .77]	.07 (.10) [–.09, .23]	.04 (.10) [–.12, .20]
MS06	–.22 (.06) [–.30, –.13]	.42 (.07) [.31, .54]	–.42 (.16) [–.65, –.16]	–1.24 (.15) [–1.48, –.99]	.85 (.22) [.53, 1.23]	.23 (.24) [–.15, .59]	.20 (.14) [.02, .44]	–.05 (.12) [–.26, .11]
MS07	–.23 (.06) [–.32, –.14]	.32 (.07) [.21, .43]	–.40 (.17) [–.64, –.10]	–1.30 (.15) [–1.54, –1.07]	.62 (.18) [.32, .89]	.46 (.23) [.12, .83]	–.04 (.12) [–.25, .12]	.08 (.11) [–.07, .27]
MS08	–.23 (.06) [–.32, –.14]	.16 (.08) [.04, .28]	–.41 (.17) [–.65, –.11]	–.81 (.15) [–1.05, –.56]	.80 (.20) [.49, 1.13]	.40 (.23) [.04, .78]	.02 (.11) [–.16, .18]	–.02 (.11) [–.22, .13]

**Table D.2 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level. ① = intercept, ② = high - (high + low)/2, ③ = Eng - (tone1 + tone4)/2, ④ = tone1 - tone4, ⑤ = asp - unasp, ⑥ = unasp - son, ⑦ = [asp - unasp] × [high - (high + low)/2], ⑧ = [unasp - son] × [high - (high + low)/2].

Par.	①	②	③	④	⑤	⑥	⑦	⑧
MS09	-.22 (.06) [-.31, -.14]	.26 (.07) [.14, .38]	-.41 (.17) [-.65, -.12]	-1.41 (.15) [-1.65, -1.16]	.56 (.20) [.22, .85]	.43 (.23) [.08, .80]	.02 (.11) [-.15, .18]	.10 (.11) [-.05, .30]
MS10	-.22 (.06) [-.31, -.14]	.28 (.07) [.18, .39]	-.41 (.16) [-.65, -.15]	-1.29 (.14) [-1.53, -1.07]	.70 (.18) [.42, .99]	.23 (.24) [-.16, .58]	.06 (.10) [-.09, .23]	-.02 (.10) [-.21, .12]
MS11	-.23 (.06) [-.31, -.14]	.40 (.07) [.29, .50]	-.41 (.17) [-.64, -.12]	-1.11 (.15) [-1.35, -.88]	.71 (.18) [.44, 1.00]	.41 (.22) [.08, .76]	.05 (.10) [-.11, .20]	.07 (.10) [-.07, .24]
MS12	-.23 (.06) [-.31, -.14]	.32 (.07) [.21, .43]	-.41 (.17) [-.64, -.14]	-1.13 (.14) [-1.35, -.90]	.70 (.18) [.42, .97]	.44 (.22) [.10, .80]	.03 (.09) [-.13, .17]	.01 (.10) [-.15, .16]
MS13	-.22 (.05) [-.31, -.14]	.45 (.07) [.34, .56]	-.42 (.16) [-.64, -.15]	-1.33 (.15) [-1.56, -1.09]	.66 (.18) [.37, .93]	.42 (.22) [.09, .76]	.07 (.10) [-.09, .23]	.02 (.10) [-.13, .17]
MS14	-.23 (.06) [-.32, -.14]	.16 (.07) [.04, .27]	-.40 (.17) [-.64, -.11]	-1.19 (.14) [-1.43, -.97]	.67 (.18) [.38, .94]	.28 (.23) [-.08, .64]	-.04 (.11) [-.24, .11]	.06 (.10) [-.09, .24]
MS15	-.23 (.06) [-.31, -.14]	.26 (.07) [.15, .36]	-.41 (.17) [-.65, -.12]	-.95 (.15) [-1.20, -.71]	.67 (.18) [.38, .94]	.38 (.22) [.03, .73]	.09 (.10) [-.06, .27]	-.02 (.11) [-.20, .13]
MS16	-.23 (.06) [-.31, -.14]	.28 (.07) [.16, .40]	-.41 (.17) [-.64, -.13]	-.85 (.15) [-1.09, -.60]	.76 (.19) [.48, 1.07]	.30 (.22) [-.07, .63]	.05 (.10) [-.11, .21]	-.02 (.11) [-.21, .13]

**Table D.2 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level. ① = intercept, ② = high - (high + low)/2, ③ = Eng - (tone1 + tone4)/2, ④ = tone1 - tone4, ⑤ = asp - unasp, ⑥ = unasp - son, ⑦ = [asp - unasp] × [high - (high + low)/2], ⑧ = [unasp - son] × [high - (high + low)/2].

258

Par.	①	②	③	④	⑤	⑥	⑦	⑧
MS17	-.23 (.06) [-.31, -.14]	.26 (.07) [.15, .37]	-.41 (.17) [-.65, -.12]	-1.46 (.15) [-1.69, -1.22]	.55 (.20) [.22, .85]	.24 (.24) [-.15, .60]	-.05 (.12) [-.26, .11]	.05 (.10) [-.11, .22]
MS18	-.23 (.06) [-.31, -.14]	.30 (.07) [.18, .41]	-.41 (.17) [-.64, -.13]	-1.37 (.15) [-1.60, -1.13]	.59 (.19) [.27, .87]	.36 (.21) [.03, .71]	.02 (.10) [-.14, .18]	.04 (.10) [-.11, .20]
MS19	-.22 (.06) [-.30, -.13]	.60 (.07) [.49, .72]	-.43 (.17) [-.68, -.16]	-.62 (.15) [-.86, -.37]	.68 (.19) [.39, .96]	.72 (.33) [.22, 1.23]	.09 (.11) [-.08, .27]	.11 (.11) [-.05, .32]
MS20	-.21 (.06) [-.30, -.12]	.60 (.07) [.48, .72]	-.45 (.16) [-.69, -.20]	-.70 (.15) [-.95, -.46]	.84 (.22) [.51, 1.19]	.37 (.23) [.00, .74]	.18 (.13) [.00, .41]	.01 (.11) [-.17, .18]
MS21	-.23 (.06) [-.31, -.14]	.33 (.07) [.22, .44]	-.41 (.16) [-.64, -.13]	-1.37 (.14) [-1.60, -1.15]	.62 (.18) [.33, .89]	.38 (.21) [.05, .71]	.08 (.10) [-.07, .26]	-.02 (.11) [-.21, .12]
MS22	-.22 (.06) [-.31, -.13]	.44 (.07) [.33, .55]	-.42 (.17) [-.66, -.14]	-.55 (.15) [-.79, -.30]	.63 (.19) [.30, .91]	.70 (.32) [.22, 1.21]	-.06 (.12) [-.28, .11]	.12 (.12) [-.04, .34]
MS23	-.23 (.06) [-.31, -.14]	.37 (.07) [.26, .48]	-.41 (.17) [-.64, -.12]	-.99 (.15) [-1.22, -.76]	.69 (.18) [.42, .96]	.59 (.26) [.20, 1.00]	.06 (.10) [-.10, .21]	.08 (.10) [-.07, .25]
MS24	-.22 (.05) [-.30, -.13]	.37 (.07) [.26, .48]	-.43 (.16) [-.66, -.17]	-.89 (.15) [-1.13, -.66]	.69 (.18) [.41, .97]	.47 (.23) [.11, .84]	.09 (.10) [-.06, .26]	.07 (.10) [-.07, .25]

**Table D.2 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level. ① = intercept, ② = high - (high + low)/2, ③ = Eng - (tone1 + tone4)/2, ④ = tone1 - tone4, ⑤ = asp - unasp, ⑥ = unasp - son, ⑦ = [asp - unasp] × [high - (high + low)/2], ⑧ = [unasp - son] × [high - (high + low)/2].

Par.	①	②	③	④	⑤	⑥	⑦	⑧
MS25	-.22 (.05) [-.30, -.13]	.32 (.07) [.21, .43]	-.42 (.16) [-.65, -.16]	-.86 (.15) [-1.09, -.62]	.71 (.18) [.44, 1.00]	.33 (.22) [-.03, .68]	.05 (.10) [-.10, .21]	-.02 (.10) [-.20, .12]

**Table D.3:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level. ⑨ = [asp - unasp] × [Eng - (tone1 + tone4)/2], ⑩ = [asp - unasp] × [tone1 - tone4], ⑪ = [unasp - son] × [Eng - (tone1 + tone4)/2], ⑫ = [unasp - son] × [tone1 - tone4].

Par.	⑨	⑩	⑪	⑫
MS01	.70 (.27) [.27, 1.14]	-.17 (.14) [-.40, .07]	.31 (.30) [-.16, .79]	.07 (.18) [-.18, .38]
MS02	.62 (.56) [-.30, 1.48]	-.19 (.16) [-.44, .05]	-.11 (.76) [-1.42, .95]	-.03 (.17) [-.30, .22]
MS03	.72 (.51) [-.08, 1.51]	-.19 (.14) [-.43, .02]	.16 (.62) [-.86, 1.06]	.00 (.16) [-.25, .24]
MS04	.74 (.52) [-.06, 1.58]	-.14 (.15) [-.36, .12]	.61 (.65) [-.29, 1.72]	.03 (.16) [-.21, .29]

**Table D.3 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level.  $\textcircled{9} = [\text{asp} - \text{unasp}] \times [\text{Eng} - (\text{tone1} + \text{tone4})/2]$ ,  $\textcircled{10} = [\text{asp} - \text{unasp}] \times [\text{tone1} - \text{tone4}]$ ,  $\textcircled{11} = [\text{unasp} - \text{son}] \times [\text{Eng} - (\text{tone1} + \text{tone4})/2]$ ,  $\textcircled{12} = [\text{unasp} - \text{son}] \times [\text{tone1} - \text{tone4}]$ .

Par.	$\textcircled{9}$	$\textcircled{10}$	$\textcircled{11}$	$\textcircled{12}$
MS05	.29 (.62) [-.78, 1.16]	-.23 (.16) [-.53, -.01]	.18 (.65) [-.89, 1.15]	.00 (.16) [-.26, .25]
MS06	.21 (.66) [-.96, 1.11]	-.22 (.16) [-.50, .01]	.58 (.67) [-.35, 1.73]	.02 (.17) [-.25, .29]
MS07	.85 (.55) [.04, 1.76]	-.17 (.15) [-.40, .06]	.10 (.65) [-1.01, 1.05]	-.06 (.18) [-.39, .18]
MS08	.38 (.60) [-.67, 1.21]	-.17 (.16) [-.42, .09]	.20 (.66) [-.85, 1.22]	.03 (.18) [-.23, .31]
MS09	1.03 (.59) [.21, 2.03]	-.12 (.16) [-.35, .15]	.20 (.64) [-.85, 1.12]	.01 (.16) [-.25, .26]
MS10	.65 (.53) [-.21, 1.43]	-.32 (.20) [-.70, -.08]	.61 (.66) [-.32, 1.75]	-.03 (.17) [-.30, .21]
MS11	.61 (.52) [-.26, 1.41]	-.20 (.14) [-.43, .01]	.21 (.62) [-.78, 1.16]	.00 (.16) [-.25, .24]
MS12	.65 (.52) [-.19, 1.44]	-.14 (.15) [-.36, .11]	.16 (.62) [-.86, 1.07]	-.01 (.16) [-.25, .23]
MS13	.74 (.52)	-.14 (.15)	.20 (.62)	-.03 (.17)

**Table D.3 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level.  $\textcircled{9} = [\text{asp} - \text{unasp}] \times [\text{Eng} - (\text{tone1} + \text{tone4})/2]$ ,  $\textcircled{10} = [\text{asp} - \text{unasp}] \times [\text{tone1} - \text{tone4}]$ ,  $\textcircled{11} = [\text{unasp} - \text{son}] \times [\text{Eng} - (\text{tone1} + \text{tone4})/2]$ ,  $\textcircled{12} = [\text{unasp} - \text{son}] \times [\text{tone1} - \text{tone4}]$ .

Par.	$\textcircled{9}$	$\textcircled{10}$	$\textcircled{11}$	$\textcircled{12}$
	[-.07, 1.58]	[-.35, .11]	[-.82, 1.11]	[-.30, .22]
MS14	.73 (.52)	-.16 (.15)	.49 (.66)	.02 (.17)
	[-.09, 1.57]	[-.39, .08]	[-.45, 1.60]	[-.22, .29]
MS15	.73 (.53)	-.18 (.15)	.29 (.62)	.03 (.17)
	[-.08, 1.59]	[-.42, .04]	[-.69, 1.27]	[-.21, .30]
MS16	.47 (.56)	-.19 (.15)	.44 (.63)	.03 (.17)
	[-.50, 1.27]	[-.43, .03]	[-.50, 1.50]	[-.21, .30]
MS17	1.04 (.60)	-.10 (.17)	.59 (.68)	.03 (.17)
	[.21, 2.06]	[-.33, .21]	[-.34, 1.73]	[-.23, .31]
MS18	.95 (.54)	-.12 (.15)	.30 (.60)	-.02 (.16)
	[.17, 1.88]	[-.33, .16]	[-.65, 1.23]	[-.28, .22]
MS19	.69 (.56)	-.13 (.17)	-.48 (.95)	-.01 (.18)
	[-.17, 1.56]	[-.38, .14]	[-2.18, .76]	[-.30, .26]
MS20	.28 (.64)	-.16 (.16)	.27 (.66)	.04 (.18)
	[-.86, 1.16]	[-.41, .10]	[-.78, 1.32]	[-.23, .35]
MS21	.86 (.51)	-.14 (.15)	.29 (.59)	.00 (.16)
	[.10, 1.72]	[-.36, .12]	[-.65, 1.23]	[-.25, .24]

**Table D.3 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M6 at the individual level. ⑨ = [asp - unasp] × [Eng - (tone1 + tone4)/2], ⑩ = [asp - unasp] × [tone1 - tone4], ⑪ = [unasp - son] × [Eng - (tone1 + tone4)/2], ⑫ = [unasp - son] × [tone1 - tone4].

Par.	⑨	⑩	⑪	⑫
MS22	.85 (.58) [-.01, 1.81]	-.17 (.16) [-.42, .08]	-.43 (.92) [-2.09, .81]	.01 (.17) [-.25, .29]
MS23	.66 (.54) [-.21, 1.48]	-.19 (.15) [-.43, .03]	-.17 (.75) [-1.52, .83]	-.05 (.17) [-.35, .18]
MS24	.68 (.52) [-.14, 1.48]	-.28 (.18) [-.62, -.06]	.08 (.66) [-1.01, 1.03]	-.01 (.16) [-.27, .23]
MS25	.61 (.53) [-.27, 1.40]	-.23 (.16) [-.50, -.01]	.38 (.62) [-.55, 1.39]	.02 (.16) [-.21, .29]

**Table D.4:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], of production VOT and post-stop F0 weights in Mandarin and English for individual speakers.

<b>Par.</b>	<b>Man. VOT wt.</b>	<b>Man. F0 wt.</b>	<b>Eng. VOT wt.</b>	<b>Eng. F0 wt.</b>
MS01	8.26 (.84) [6.92, 9.64]	.53 (.15) [.29, .77]	5.77 (.70) [4.66, 6.93]	1.51 (.35) [.97, 2.10]
MS02	7.45 (.76) [6.26, 8.66]	.65 (.17) [.38, .93]	4.64 (.51) [3.84, 5.46]	.52 (.24) [.12, .89]
MS03	6.73 (.70) [5.64, 7.88]	.35 (.16) [.11, .61]	4.78 (.64) [3.84, 5.87]	.29 (.29) [−.19, .74]
MS04	6.74 (.73) [5.60, 7.94]	.32 (.17) [.06, .61]	4.81 (.60) [3.89, 5.84]	.69 (.35) [.13, 1.24]
MS05	7.47 (.72) [6.33, 8.64]	.54 (.16) [.31, .80]	6.43 (.83) [5.12, 7.79]	.71 (.26) [.28, 1.14]
MS06	7.70 (.80) [6.46, 8.99]	.84 (.20) [.53, 1.17]	6.51 (.86) [5.22, 7.94]	.83 (.28) [.39, 1.27]
MS07	5.89 (.59) [4.96, 6.85]	.44 (.14) [.21, .66]	4.47 (.50) [3.70, 5.27]	.79 (.26) [.39, 1.22]
MS08	4.49 (.51) [3.68, 5.34]	.75 (.21) [.44, 1.09]	2.86 (.42) [2.20, 3.54]	.58 (.24) [.20, .98]
MS09	7.67 (.82) [6.37, 8.99]	.22 (.19) [−.10, .51]	5.22 (.63) [4.21, 6.23]	1.23 (.37) [.66, 1.84]
MS10	5.35 (.53) [4.52, 6.22]	.48 (.13) [.28, .69]	4.28 (.52) [3.48, 5.12]	.87 (.28) [.43, 1.31]
MS11	5.83 (.62) [4.89, 6.84]	.45 (.15) [.22, .69]	4.47 (.52) [3.66, 5.30]	.91 (.28) [.48, 1.37]
MS12	4.57 (.45) [3.86, 5.29]	.51 (.14) [.29, .74]	4.93 (.70) [3.89, 6.15]	1.00 (.31) [.51, 1.51]
MS13	6.66 (.67) [5.62, 7.77]	.29 (.17) [.01, .55]	5.73 (.69) [4.65, 6.83]	.25 (.29) [−.21, .71]
MS14	10.77 (1.21)	.29 (.18)	6.81 (.78)	.87 (.32)

**Table D.4 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], of production VOT and post-stop F0 weights in Mandarin and English for individual speakers.

<b>Par.</b>	<b>Man. VOT wt.</b>	<b>Man. F0 wt.</b>	<b>Eng. VOT wt.</b>	<b>Eng. F0 wt.</b>
	[8.90, 12.77]	[.00, .60]	[5.60, 8.08]	[.36, 1.38]
MS15	7.14 (.79)	.31 (.15)	4.80 (.64)	.59 (.31)
	[5.91, 8.44]	[.07, .54]	[3.76, 5.82]	[.09, 1.10]
MS16	7.91 (.83)	.55 (.17)	5.13 (.62)	1.21 (.32)
	[6.62, 9.26]	[.29, .83]	[4.18, 6.15]	[.73, 1.75]
MS17	6.20 (.66)	.25 (.16)	4.84 (.66)	1.38 (.35)
	[5.18, 7.26]	[.00, .50]	[3.82, 5.94]	[.83, 1.95]
MS18	5.94 (.67)	.38 (.13)	4.16 (.61)	.84 (.30)
	[4.89, 7.02]	[.16, .58]	[3.21, 5.17]	[.35, 1.33]
MS19	8.78 (.92)	.44 (.15)	7.34 (1.05)	.33 (.25)
	[7.36, 10.26]	[.20, .68]	[5.64, 9.02]	[-.09, .71]
MS20	5.64 (.60)	.47 (.18)	4.57 (.66)	.71 (.40)
	[4.71, 6.61]	[.21, .78]	[3.59, 5.66]	[.13, 1.38]
MS21	5.78 (.61)	.36 (.15)	4.98 (.61)	1.91 (.41)
	[4.83, 6.81]	[.12, .59]	[4.03, 5.97]	[1.25, 2.58]
MS22	8.51 (.85)	.33 (.13)	5.95 (.71)	.63 (.31)
	[7.18, 9.88]	[.12, .54]	[4.86, 7.13]	[.12, 1.13]
MS23	7.27 (.78)	.50 (.15)	6.10 (.76)	1.09 (.29)
	[6.04, 8.53]	[.27, .74]	[4.93, 7.35]	[.64, 1.58]
MS24	6.29 (.63)	.38 (.12)	5.24 (.68)	1.22 (.35)
	[5.28, 7.30]	[.19, .57]	[4.20, 6.40]	[.67, 1.78]
MS25	7.87 (.84)	.44 (.15)	5.08 (.62)	.84 (.26)
	[6.56, 9.20]	[.20, .68]	[4.07, 6.05]	[.43, 1.27]

**Table D.5:** Summary of posterior distributions for guessing probabilities, in terms of mean (sd) [89% CrI].

<b>Par.</b>	<b>Guessing prob. (Man.)</b>	<b>Par.</b>	<b>Guessing prob. (Eng.)</b>
MS01	.01 (.01) [.00, .02]	MS01	.01 (.01) [.00, .02]
MS02	.01 (.01) [.00, .02]	MS02	.01 (.01) [.00, .02]
MS03	.01 (.01) [.00, .02]	MS03	.01 (.01) [.00, .02]
MS04	.01 (.01) [.00, .03]	MS04	.04 (.02) [.01, .07]
MS05	.02 (.01) [.01, .05]	MS05	.01 (.01) [.00, .03]
MS06	.02 (.01) [.00, .04]	MS06	.01 (.01) [.00, .02]
MS07	.02 (.01) [.00, .04]	MS07	.01 (.01) [.00, .04]
MS08	.02 (.01) [.00, .04]	MS08	.01 (.01) [.00, .04]
MS09	.01 (.01) [.00, .02]	MS09	.01 (.01) [.00, .02]
MS10	.02 (.01) [.00, .05]	MS10	.01 (.01) [.00, .02]
MS11	.01 (.01) [.00, .03]	MS11	.01 (.01) [.00, .02]
MS12	.02 (.01) [.01, .05]	MS12	.09 (.02) [.05, .13]
MS13	.01 (.01) [.00, .02]	MS13	.01 (.01) [.00, .02]
MS14	.02 (.01) [.00, .04]	MS14	.01 (.01) [.00, .02]
MS15	.01 (.01) [.00, .02]	MS15	.03 (.01) [.01, .06]
MS16	.01 (.01) [.00, .02]	MS16	.01 (.01) [.00, .03]
MS17	.01 (.01) [.00, .02]	MS17	.01 (.01) [.00, .03]
MS18	.01 (.01) [.00, .02]	MS18	.01 (.01) [.00, .03]
MS19	.01 (.01) [.00, .02]	MS19	.01 (.01) [.00, .03]
MS20	.01 (.01) [.00, .02]	MS20	.02 (.01) [.00, .04]
MS21	.02 (.02) [.00, .05]	MS21	.01 (.01) [.00, .02]
MS22	.01 (.01) [.00, .02]	MS22	.03 (.01) [.01, .05]
MS23	.02 (.01) [.00, .04]	MS23	.01 (.01) [.00, .03]
MS24	.01 (.01) [.00, .02]	MS24	.01 (.01) [.00, .02]
MS25	.03 (.02) [.01, .05]	MS25	.01 (.01) [.00, .03]

**Table D.6:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M4 at the individual level.

<b>Par.</b>	<b>intercept<sub>Man</sub></b>	<b>VOT<sub>Man</sub></b>	<b>F0<sub>Man</sub></b>	<b>tone<sub>Man</sub></b>	<b>intercept<sub>Eng</sub></b>	<b>VOT<sub>Eng</sub></b>	<b>F0<sub>Eng</sub></b>	<b>tone<sub>Eng</sub></b>
MS01	10.49 (1.65) [8.16, 13.30]	13.39 (1.94) [10.61, 16.68]	.89 (.32) [.42, 1.44]	.70 (.26) [.38, 1.16]	10.19 (1.45) [8.06, 12.68]	14.25 (1.92) [11.50, 17.56]	1.39 (.45) [.79, 2.19]	.48 (.25) [.10, .90]
MS02	9.58 (1.31) [7.66, 11.75]	15.97 (2.46) [12.50, 20.17]	.15 (.35) [−.45, .65]	.54 (.23) [.19, .91]	11.20 (1.56) [8.97, 13.89]	16.54 (2.21) [13.35, 20.37]	.65 (.40) [−.01, 1.24]	.34 (.28) [−.13, .76]
MS03	11.12 (1.60) [8.84, 13.90]	17.18 (2.57) [13.44, 21.62]	.34 (.34) [−.23, .85]	.49 (.24) [.11, .86]	10.48 (1.28) [8.56, 12.64]	17.27 (2.32) [13.81, 21.27]	.80 (.35) [.22, 1.36]	.25 (.27) [−.22, .63]
MS04	10.33 (1.57) [8.09, 13.03]	15.44 (2.41) [12.03, 19.60]	.64 (.38) [.05, 1.26]	.44 (.28) [−.03, .84]	7.49 (1.77) [4.61, 10.32]	14.99 (3.74) [9.22, 21.10]	1.11 (.37) [.61, 1.77]	.25 (.28) [−.23, .64]
MS05	8.53 (1.40) [6.50, 10.82]	12.55 (1.90) [9.70, 15.70]	.67 (.29) [.23, 1.15]	.60 (.22) [.29, .96]	9.84 (1.25) [8.00, 11.90]	14.97 (1.92) [12.18, 18.14]	.87 (.32) [.35, 1.38]	.39 (.25) [−.01, .75]
MS06	8.61 (1.23) [6.75, 10.64]	14.91 (2.37) [11.48, 18.84]	.53 (.29) [.09, 1.01]	.51 (.21) [.19, .84]	9.65 (1.22) [7.83, 11.65]	15.34 (2.05) [12.36, 18.84]	.76 (.32) [.23, 1.23]	.41 (.25) [.01, .80]
MS07	6.87 (1.17) [5.11, 8.83]	11.25 (2.11) [8.26, 14.85]	.75 (.29) [.32, 1.23]	.49 (.22) [.13, .81]	8.83 (1.51) [6.64, 11.38]	11.79 (1.86) [9.03, 14.94]	.90 (.31) [.42, 1.43]	.46 (.25) [.07, .84]
MS08	8.69 (1.49) [6.52, 11.22]	13.95 (2.63) [10.37, 18.58]	.28 (.34) [−.29, .76]	.46 (.24) [.05, .81]	8.46 (1.28) [6.47, 10.55]	15.67 (2.63) [11.84, 20.25]	.54 (.33) [−.01, 1.02]	.26 (.26) [−.19, .62]
MS09	6.99 (1.02)	10.70 (1.48)	.76 (.27)	.67 (.23)	9.36 (1.19)	13.52 (1.63)	1.07 (.32)	.62 (.25)

**Table D.6 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M4 at the individual level.

Par.	intercept <sub>Man</sub>	VOT <sub>Man</sub>	F0 <sub>Man</sub>	tone <sub>Man</sub>	intercept <sub>Eng</sub>	VOT <sub>Eng</sub>	F0 <sub>Eng</sub>	tone <sub>Eng</sub>
	[5.43, 8.70]	[8.40, 13.22]	[.35, 1.21]	[.37, 1.07]	[7.62, 11.33]	[11.04, 16.29]	[.61, 1.64]	[.28, 1.05]
MS10	6.40 (1.35)	12.47 (2.79)	.85 (.32)	.51 (.21)	8.32 (1.25)	15.28 (2.46)	1.10 (.35)	.76 (.31)
	[4.45, 8.74]	[8.64, 17.37]	[.38, 1.40]	[.17, .85]	[6.54, 10.43]	[11.80, 19.42]	[.62, 1.72]	[.35, 1.31]
MS11	11.02 (1.63)	15.34 (2.11)	.43 (.32)	.51 (.23)	10.97 (1.49)	16.23 (2.08)	.92 (.37)	.39 (.27)
	[8.72, 13.84]	[12.34, 18.92]	[−.09, .92]	[.14, .87]	[8.90, 13.53]	[13.27, 19.74]	[.35, 1.54]	[−.02, .80]
MS12	8.90 (1.35)	16.28 (2.76)	.18 (.33)	.33 (.25)	8.88 (1.30)	16.77 (2.65)	.54 (.34)	.36 (.24)
	[6.89, 11.18]	[12.35, 21.07]	[−.36, .65]	[−.11, .67]	[6.91, 11.06]	[12.77, 21.30]	[−.02, 1.03]	[−.03, .73]
MS13	11.27 (1.75)	15.47 (2.18)	.47 (.33)	.53 (.23)	11.67 (1.63)	15.78 (1.98)	.94 (.34)	.32 (.26)
	[8.81, 14.22]	[12.35, 19.16]	[−.07, .98]	[.16, .88]	[9.37, 14.53]	[12.96, 19.18]	[.41, 1.51]	[−.12, .70]
MS14	11.60 (1.80)	15.94 (2.34)	.34 (.35)	.52 (.25)	11.95 (1.64)	16.49 (2.05)	.98 (.37)	.17 (.31)
	[9.10, 14.73]	[12.60, 19.87]	[−.24, .85]	[.12, .90]	[9.61, 14.75]	[13.48, 19.98]	[.45, 1.61]	[−.39, .59]
MS15	11.20 (1.61)	17.19 (2.57)	.32 (.34)	.51 (.24)	10.94 (1.44)	17.10 (2.38)	.73 (.37)	.27 (.29)
	[8.87, 13.90]	[13.52, 21.62]	[−.26, .82]	[.13, .87]	[8.87, 13.36]	[13.62, 21.17]	[.11, 1.25]	[−.23, .67]
MS16	9.13 (1.29)	15.58 (2.51)	.22 (.32)	.55 (.22)	10.67 (1.42)	16.23 (2.16)	.64 (.37)	.45 (.27)
	[7.21, 11.36]	[11.97, 19.92]	[−.33, .68]	[.22, .90]	[8.68, 13.06]	[13.21, 19.88]	[.01, 1.18]	[.03, .89]
MS17	11.83 (1.72)	17.51 (2.52)	.31 (.36)	.50 (.24)	11.24 (1.47)	16.91 (2.23)	.87 (.36)	.20 (.31)
	[9.39, 14.82]	[13.86, 21.88]	[−.30, .83]	[.10, .87]	[9.13, 13.74]	[13.70, 20.81]	[.33, 1.45]	[−.34, .62]
MS18	10.01 (1.65)	12.61 (1.85)	.42 (.30)	.55 (.22)	11.49 (1.63)	16.17 (2.15)	.91 (.36)	.37 (.27)

**Table D.6 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M4 at the individual level.

<b>Par.</b>	<b>intercept<sub>Man</sub></b>	<b>VOT<sub>Man</sub></b>	<b>F0<sub>Man</sub></b>	<b>tone<sub>Man</sub></b>	<b>intercept<sub>Eng</sub></b>	<b>VOT<sub>Eng</sub></b>	<b>F0<sub>Eng</sub></b>	<b>tone<sub>Eng</sub></b>
	[7.68, 12.70]	[9.92, 15.67]	[−.09, .86]	[.21, .89]	[9.20, 14.34]	[13.11, 19.87]	[.34, 1.49]	[−.08, .77]
MS19	8.55 (1.28)	15.28 (2.57)	.63 (.30)	.50 (.21)	8.17 (1.28)	15.91 (2.65)	.76 (.28)	.52 (.23)
	[6.65, 10.70]	[11.63, 19.71]	[.18, 1.11]	[.17, .84]	[6.20, 10.30]	[11.92, 20.42]	[.30, 1.18]	[.18, .92]
MS20	7.07 (1.07)	8.11 (1.21)	.98 (.28)	.72 (.22)	9.92 (1.73)	13.19 (2.31)	1.12 (.35)	.61 (.25)
	[5.50, 8.84]	[6.38, 10.06]	[.56, 1.45]	[.42, 1.12]	[7.49, 12.93]	[9.85, 17.24]	[.64, 1.69]	[.25, 1.04]
MS21	7.70 (1.55)	10.01 (1.84)	.64 (.28)	.46 (.21)	9.22 (1.21)	13.78 (1.76)	1.15 (.36)	.48 (.25)
	[5.42, 10.29]	[7.27, 13.13]	[.20, 1.08]	[.10, .75]	[7.41, 11.26]	[11.12, 16.75]	[.68, 1.80]	[.10, .89]
MS22	7.89 (1.38)	15.80 (2.95)	.16 (.29)	.37 (.22)	9.48 (1.34)	16.74 (2.66)	.40 (.38)	.43 (.26)
	[5.83, 10.25]	[11.51, 20.82]	[−.31, .60]	[.00, .68]	[7.50, 11.73]	[12.88, 21.34]	[−.23, .95]	[.02, .86]
MS23	9.66 (1.88)	11.48 (2.16)	.81 (.30)	.77 (.27)	11.19 (1.74)	13.04 (1.97)	1.23 (.35)	.55 (.24)
	[6.98, 12.90]	[8.40, 15.13]	[.37, 1.31]	[.42, 1.26]	[8.61, 14.23]	[10.11, 16.45]	[.74, 1.85]	[.19, .95]
MS24	11.34 (1.58)	17.39 (2.52)	.35 (.34)	.53 (.24)	11.56 (1.54)	17.75 (2.46)	.84 (.37)	.23 (.29)
	[9.08, 14.05]	[13.76, 21.66]	[−.20, .87]	[.16, .91]	[9.36, 14.24]	[14.21, 22.10]	[.29, 1.43]	[−.29, .63]
MS25	5.28 (.93)	5.95 (1.04)	1.05 (.30)	.60 (.21)	6.63 (1.19)	7.38 (1.30)	.91 (.29)	.76 (.28)
	[3.91, 6.85]	[4.42, 7.72]	[.59, 1.55]	[.28, .95]	[4.91, 8.65]	[5.53, 9.62]	[.45, 1.38]	[.37, 1.25]

**Table D.7:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M4 at the individual level.

<b>Par.</b>	<b>inter<sub>Man</sub> – inter<sub>Eng</sub></b>	<b>VOT<sub>Man</sub> – VOT<sub>Eng</sub></b>	<b>F0<sub>Man</sub> – F0<sub>Eng</sub></b>	<b>tone<sub>Man</sub> – tone<sub>Eng</sub></b>
MS1	.30 (1.96) [–2.73, 3.55]	–.86 (2.45) [–4.71, 3.15]	–.50 (.51) [–1.37, .26]	.23 (.35) [–.31, .80]
MS2	–1.62 (1.80) [–4.59, 1.10]	–.57 (2.85) [–4.94, 4.04]	–.50 (.48) [–1.29, .24]	.20 (.36) [–.34, .79]
MS3	.64 (1.80) [–2.12, 3.67]	–.09 (2.90) [–4.46, 4.66]	–.46 (.44) [–1.18, .21]	.24 (.35) [–.28, .83]
MS4	2.85 (2.09) [–.32, 6.30]	.45 (3.85) [–5.62, 6.65]	–.48 (.48) [–1.28, .27]	.19 (.38) [–.41, .81]
MS5	–1.32 (1.67) [–3.97, 1.24]	–2.41 (2.39) [–6.23, 1.23]	–.20 (.40) [–.79, .45]	.20 (.32) [–.27, .75]
MS6	–1.03 (1.56) [–3.54, 1.43]	–.43 (2.72) [–4.57, 4.01]	–.23 (.39) [–.84, .39]	.10 (.31) [–.39, .59]
MS7	–1.96 (1.74) [–4.80, .68]	–.53 (2.57) [–4.53, 3.65]	–.16 (.39) [–.77, .47]	.04 (.32) [–.47, .54]
MS8	.24 (1.68) [–2.36, 2.98]	–1.73 (3.06) [–6.62, 3.18]	–.26 (.42) [–.92, .41]	.20 (.34) [–.31, .75]
MS9	–2.36 (1.46) [–4.73, –.07]	–2.82 (2.06) [–6.06, .39]	–.31 (.38) [–.95, .29]	.04 (.33) [–.48, .56]
MS10	–1.93 (1.66) [–4.54, .61]	–2.81 (3.24) [–7.89, 2.19]	–.25 (.43) [–.94, .41]	–.25 (.37) [–.88, .29]
MS11	.05 (1.96) [–3.03, 3.16]	–.90 (2.63) [–4.98, 3.29]	–.49 (.45) [–1.22, .19]	.12 (.35) [–.44, .66]
MS12	.02 (1.60) [–2.47, 2.63]	–.49 (3.13) [–5.40, 4.62]	–.36 (.43) [–1.07, .33]	–.03 (.34) [–.59, .49]
MS13	–.40 (2.06) [–3.64, 2.83]	–.31 (2.58) [–4.37, 3.87]	–.48 (.44) [–1.19, .18]	.21 (.34) [–.31, .76]
MS14	–.34 (2.11) [–3.66, 3.05]	–.56 (2.70) [–4.76, 3.75]	–.64 (.48) [–1.45, .06]	.35 (.39) [–.22, 1.00]

**Table D.7 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M4 at the individual level.

<b>Par.</b>	<b>inter<sub>Man</sub> – inter<sub>Eng</sub></b>	<b>VOT<sub>Man</sub> – VOT<sub>Eng</sub></b>	<b>F0<sub>Man</sub> – F0<sub>Eng</sub></b>	<b>tone<sub>Man</sub> – tone<sub>Eng</sub></b>
MS15	.26 (1.90) [–2.64, 3.34]	.09 (3.02) [–4.56, 5.13]	–.41 (.45) [–1.12, .29]	.24 (.36) [–.29, .84]
MS16	–1.54 (1.72) [–4.36, 1.08]	–.65 (2.86) [–5.00, 4.03]	–.42 (.44) [–1.13, .26]	.10 (.34) [–.43, .66]
MS17	.60 (1.97) [–2.43, 3.85]	.60 (2.84) [–3.81, 5.27]	–.57 (.45) [–1.32, .10]	.29 (.37) [–.27, .93]
MS18	–1.48 (2.00) [–4.71, 1.58]	–3.56 (2.47) [–7.63, .19]	–.49 (.45) [–1.23, .17]	.18 (.34) [–.35, .71]
MS19	.38 (1.61) [–2.06, 3.00]	–.63 (3.19) [–5.53, 4.53]	–.13 (.37) [–.70, .49]	–.02 (.31) [–.52, .46]
MS20	–2.85 (1.92) [–6.19, –.09]	–5.08 (2.47) [–9.28, –1.51]	–.15 (.42) [–.83, .50]	.11 (.33) [–.42, .64]
MS21	–1.52 (1.78) [–4.42, 1.28]	–3.77 (2.31) [–7.49, –.09]	–.51 (.42) [–1.22, .10]	–.02 (.31) [–.53, .45]
MS22	–1.59 (1.65) [–4.30, 1.02]	–.94 (3.27) [–6.13, 4.30]	–.24 (.44) [–.94, .48]	–.06 (.33) [–.61, .45]
MS23	–1.54 (2.24) [–5.06, 1.97]	–1.56 (2.57) [–5.62, 2.43]	–.43 (.43) [–1.13, .23]	.22 (.36) [–.33, .83]
MS24	–.23 (1.87) [–3.19, 2.70]	–.36 (2.93) [–4.91, 4.26]	–.50 (.46) [–1.21, .20]	.30 (.37) [–.24, .94]
MS25	–1.35 (1.45) [–3.74, .89]	–1.43 (1.61) [–4.04, 1.04]	.13 (.40) [–.51, .79]	–.17 (.34) [–.72, .35]

## Appendix E

# Supporting Materials for Chapter 4

### E.1 Participant Demographic Information

**Table E.1:** Demographic information of the L1 English participants. Am. = American English; Br. = British English; Ca. = Canadian English.

ID	Gender	Age	Eng. variety	Other lang.
1	F	19	Ca.	German
2	F	19	Ca.	
3	F	20	Ca.	French
4	F	19	Am./Br./Ca.	
5	F	20	Ca.	
6	M	44	Ca.	Japanese
7	F	23	Ca.	
8	M	24	Ca.	Punjabi
9	M	19	Ca.	French
10	F	33	Am.	
11	F	30	Ca.	
12	F	29	Ca.	Portuguese

**Table E.1 continued:** Demographic information of the L1 English participants. Am. = American English; Br. = British English; Ca. = Canadian English.

ID	Gender	Age	Eng. variety	Other lang.
13	F	23	Am./Ca.	French
14	F	22	Am./Ca.	
15	M	46	Ca.	French/German
16	F	19	Am./Ca.	French
17	F	24	Ca.	
18	F	19	Ca.	French
19	M	27	Ca.	
20	F	20	Am.	
21	F	47	Ca.	
22	M	20	Am./Ca./AAVE	Russian/Toki Pona
23	F	21	Ca.	
24	M	21	Ca.	Greek
25	M	25	Ca.	French

## E.2 Statistical Model Specification

### E.2.1 Production Model for Post-Stop F0

This model concerns the English productions from L1 English speakers and L1 Mandarin-L2 English speakers. In the model, the coefficients are decomposed into three parts: a language-agnostic estimate (e.g.,  $\beta_0$ ), a speaker-group-specific adjustment (e.g.,  ${}^\ell w_0$ ), and a participant-specific adjustment within each language (e.g.,  ${}^\ell u_0^j$ ).

Likelihood:

$$\begin{aligned} {}^{\ell}\text{FO}_i^j &= (\beta_0 + {}^{\ell}w_0 + {}^{\ell}u_0^j) + && \\ &(\beta_1 + {}^{\ell}w_1 + {}^{\ell}u_1^j) \cdot \text{HIGH}_i + && \text{(vowel height)} \\ &(\beta_2 + {}^{\ell}w_2 + {}^{\ell}u_2^j) \cdot [\text{ASP vs. UNASP}]_i + \\ &(\beta_3 + {}^{\ell}w_3 + {}^{\ell}u_3^j) \cdot [\text{UNASP vs. SON}]_i + && \text{(voicing)} \\ &\varepsilon && \text{(error term)} \\ \varepsilon &\sim \mathcal{N}(0, \sigma^2) \end{aligned}$$

Priors:

$$\beta_0, \beta_1, \dots, \beta_3 \sim \mathcal{N}(0, 5)$$

$$\sigma \sim \text{Exponential}(1)$$

$$\begin{bmatrix} \ell w_0 \\ \ell w_1 \\ \vdots \\ \ell w_3 \end{bmatrix} \sim \mathcal{N}(0, S), \ell = \text{Eng}, \text{Man}$$

$$\begin{bmatrix} \ell=\text{Eng} \mu_0^j \\ \ell=\text{Eng} \mu_1^j \\ \vdots \\ \ell=\text{Eng} \mu_3^j \end{bmatrix} \sim \mathcal{N}(0, \ell=\text{Eng} S), j = 1, 2, \dots, 25$$

$$\begin{bmatrix} \ell=\text{Man} \mu_0^j \\ \ell=\text{Man} \mu_1^j \\ \vdots \\ \ell=\text{Man} \mu_3^j \end{bmatrix} \sim \mathcal{N}(0, \ell=\text{Man} S), j = 1, 2, \dots, 25$$

$$S = \begin{bmatrix} \sigma_0 & 0 & \cdots & 0 \\ 0 & \sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \sigma_3 \end{bmatrix} R = \begin{bmatrix} \sigma_0 & 0 & \cdots & 0 \\ 0 & \sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \sigma_3 \end{bmatrix}$$

$$\ell=\text{Eng} S = \begin{bmatrix} \ell=\text{Eng} \sigma_0 & 0 & \cdots & 0 \\ 0 & \ell=\text{Eng} \sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \ell=\text{Eng} \sigma_3 \end{bmatrix} \ell=\text{Eng} R = \begin{bmatrix} \ell=\text{Eng} \sigma_0 & 0 & \cdots & 0 \\ 0 & \ell=\text{Eng} \sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \ell=\text{Eng} \sigma_3 \end{bmatrix}$$

$$\ell=\text{Man} S = \begin{bmatrix} \ell=\text{Man} \sigma_0 & 0 & \cdots & 0 \\ 0 & \ell=\text{Man} \sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \ell=\text{Man} \sigma_3 \end{bmatrix} \ell=\text{Man} R = \begin{bmatrix} \ell=\text{Man} \sigma_0 & 0 & \cdots & 0 \\ 0 & \ell=\text{Man} \sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \ell=\text{Man} \sigma_3 \end{bmatrix}$$

$$\sigma_0, \sigma_1, \dots, \sigma_3 \sim \text{Exponential}(1)$$

$$\ell=\text{Eng} \sigma_0, \ell=\text{Eng} \sigma_1, \dots, \ell=\text{Eng} \sigma_3 \sim \text{Exponential}(1)$$

$$\ell=\text{Man} \sigma_0, \ell=\text{Man} \sigma_1, \dots, \ell=\text{Man} \sigma_3 \sim \text{Exponential}(1)$$

$$R \sim \text{LKJCorr}(1)$$

$$\ell=\text{Eng} R \sim \text{LKJCorr}(1)$$

$$\ell=\text{Man} R \sim \text{LKJCorr}(1)$$

### E.2.2 Perceptual Model for Post-Stop F0 Weight

Similar to the model in Section E.2.1, each coefficient is decomposed into three parts that consist of a language-universal estimate, a listener-group-specific adjustment, and a participant-specific adjustment. In addition, the model assumes that, for each trial, there is a certain chance  ${}^\ell P_{\text{guess}}^j$ , to be estimated in the model, that a particular response was registered due to random selection (i.e., 0.25 in this study because there are four options in each trial) because the participant was distracted or simply made a mistake, and that for a chance of  $(1 - {}^\ell P_{\text{guess}}^j)$  the response was chosen according to the probability output by the logistic function.

Likelihood:

$$\begin{aligned}
 {}^\ell \text{/p/ response}_i^j \sim \text{Bernoulli}(& \\
 & {}^\ell P_{\text{guess}}^j \cdot 0.25 + \\
 & (1 - {}^\ell P_{\text{guess}}^j) \cdot \text{logit}^{-1}(& \\
 & (\beta_0 + {}^\ell w_0 + {}^\ell u_0^j) + & \\
 & (\beta_1 + {}^\ell w_1 + {}^\ell u_1^j) \cdot \text{VOT}_i + & \text{(VOT)} \\
 & (\beta_2 + {}^\ell w_2 + {}^\ell u_2^j) \cdot \text{F0}_i + & \text{(F0)} \\
 & (\beta_3 + {}^\ell w_3 + {}^\ell u_3^j) \cdot \text{TONE } 1_i & \text{(tone)} \\
 & )) & 
 \end{aligned}$$

Priors:

$${}^{\ell=\text{Eng}}p_{\text{guess}}^j, {}^{\ell=\text{Man}}p_{\text{guess}}^j \sim \text{Unif}(0, 1), j = 1, 2, \dots, 25$$

$$\beta_0, \beta_1, \dots, \beta_3 \sim \mathcal{N}(0, 10)$$

$$\begin{bmatrix} {}^{\ell}w_0 \\ {}^{\ell}w_1 \\ \vdots \\ {}^{\ell}w_3 \end{bmatrix} \sim \mathcal{N}(0, S), \ell = \text{Eng}, \text{Man}$$

$$\begin{bmatrix} {}^{\ell=\text{Eng}}\mu_0^j \\ {}^{\ell=\text{Eng}}\mu_1^j \\ \vdots \\ {}^{\ell=\text{Eng}}\mu_3^j \end{bmatrix} \sim \mathcal{N}(0, {}^{\ell=\text{Eng}}S), j = 1, 2, \dots, 25$$

$$\begin{bmatrix} {}^{\ell=\text{Man}}\mu_0^j \\ {}^{\ell=\text{Man}}\mu_1^j \\ \vdots \\ {}^{\ell=\text{Man}}\mu_3^j \end{bmatrix} \sim \mathcal{N}(0, {}^{\ell=\text{Man}}S), j = 1, 2, \dots, 25$$

$$S = \begin{bmatrix} \sigma_0 & 0 & \cdots & 0 \\ 0 & \sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \sigma_3 \end{bmatrix} R = \begin{bmatrix} \sigma_0 & 0 & \cdots & 0 \\ 0 & \sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \sigma_3 \end{bmatrix}$$

$${}^{\ell=\text{Eng}}S = \begin{bmatrix} {}^{\ell=\text{Eng}}\sigma_0 & 0 & \cdots & 0 \\ 0 & {}^{\ell=\text{Eng}}\sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & {}^{\ell=\text{Eng}}\sigma_3 \end{bmatrix} {}^{\ell=\text{Eng}}R = \begin{bmatrix} {}^{\ell=\text{Eng}}\sigma_0 & 0 & \cdots & 0 \\ 0 & {}^{\ell=\text{Eng}}\sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & {}^{\ell=\text{Eng}}\sigma_3 \end{bmatrix}$$

$${}^{\ell=\text{Man}}S = \begin{bmatrix} {}^{\ell=\text{Man}}\sigma_0 & 0 & \cdots & 0 \\ 0 & {}^{\ell=\text{Man}}\sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & {}^{\ell=\text{Man}}\sigma_3 \end{bmatrix} {}^{\ell=\text{Man}}R = \begin{bmatrix} {}^{\ell=\text{Man}}\sigma_0 & 0 & \cdots & 0 \\ 0 & {}^{\ell=\text{Man}}\sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & {}^{\ell=\text{Man}}\sigma_3 \end{bmatrix}$$

$$\sigma_0, \sigma_1, \dots, \sigma_3 \sim \text{Exponential}(1)$$

$${}^{\ell=\text{Eng}}\sigma_0, {}^{\ell=\text{Eng}}\sigma_1, \dots, {}^{\ell=\text{Eng}}\sigma_3 \sim \text{Exponential}(1)$$

$${}^{\ell=\text{Man}}\sigma_0, {}^{\ell=\text{Man}}\sigma_1, \dots, {}^{\ell=\text{Man}}\sigma_3 \sim \text{Exponential}(1)$$

$$R \sim \text{LKJCorr}(1)$$

$${}^{\ell=\text{Eng}}R \sim \text{LKJCorr}(1)$$

$${}^{\ell=\text{Man}}R \sim \text{LKJCorr}(1)$$

### E.3 Individual-Level Posterior Parameter Summaries

**Table E.2:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M2 for L1 English speakers at the individual level.

Par.	intercept	high – (high + low)/2	asp – unasp	unasp – son
E1	–.01 (.03) [–.05, .03]	.31 (.07) [.20, .43]	.80 (.20) [.47, 1.13]	.77 (.21) [.45, 1.11]
E2	.00 (.03) [–.04, .04]	.27 (.07) [.16, .39]	1.12 (.21) [.79, 1.45]	.49 (.20) [.17, .82]
E3	–.01 (.03) [–.04, .03]	.30 (.07) [.18, .42]	.95 (.22) [.59, 1.32]	.50 (.23) [.14, .87]
E4	–.01 (.03) [–.05, .03]	.28 (.07) [.17, .39]	.57 (.20) [.26, .90]	.78 (.21) [.45, 1.11]
E5	–.01 (.03) [–.05, .04]	.25 (.07) [.13, .36]	.63 (.21) [.30, .95]	1.03 (.21) [.71, 1.38]
E6	.00 (.03) [–.04, .05]	.18 (.08) [.03, .30]	1.15 (.22) [.79, 1.50]	.26 (.21) [–.09, .59]
E7	–.01 (.03) [–.05, .03]	.37 (.07) [.26, .49]	.94 (.21) [.60, 1.28]	.13 (.21) [–.20, .47]
E8	–.01 (.03) [–.05, .03]	.26 (.07) [.14, .37]	.66 (.21) [.33, .98]	.79 (.21) [.47, 1.12]
E9	–.01 (.03) [–.05, .03]	.33 (.07) [.23, .45]	1.12 (.21) [.80, 1.47]	.22 (.21) [–.12, .54]
E10	–.01 (.03) [–.05, .03]	.29 (.07) [.17, .40]	.28 (.21) [–.05, .61]	.86 (.21) [.53, 1.19]
E11	.00 (.03) [–.04, .04]	.31 (.07) [.20, .43]	1.24 (.21) [.91, 1.58]	.16 (.21) [–.18, .50]
E12	–.01 (.03) [–.05, .04]	.44 (.08) [.32, .57]	1.29 (.21) [.96, 1.63]	–.08 (.21) [–.41, .26]

**Table E.2 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M2 for L1 English speakers at the individual level.

<b>Par.</b>	<b>intercept</b>	<b>high – (high + low)/2</b>	<b>asp – unasp</b>	<b>unasp – son</b>
E13	–.01 (.03) [–.05, .03]	.32 (.07) [.21, .44]	1.23 (.21) [.90, 1.58]	.08 (.21) [–.26, .41]
E14	–.01 (.03) [–.06, .03]	.42 (.08) [.30, .55]	.63 (.21) [.29, .96]	.26 (.22) [–.09, .60]
E15	.00 (.03) [–.05, .04]	.24 (.07) [.12, .36]	1.47 (.21) [1.13, 1.81]	.29 (.21) [–.04, .63]
E16	–.01 (.03) [–.05, .04]	.22 (.07) [.10, .33]	.35 (.21) [.01, .68]	1.13 (.21) [.79, 1.46]
E17	–.01 (.03) [–.05, .03]	.34 (.07) [.22, .45]	.81 (.21) [.48, 1.14]	.26 (.21) [–.08, .60]
E18	–.01 (.03) [–.06, .03]	.37 (.07) [.26, .49]	.31 (.21) [–.03, .64]	.70 (.21) [.36, 1.03]
E19	–.01 (.03) [–.05, .03]	.41 (.07) [.30, .53]	.95 (.20) [.63, 1.28]	.06 (.20) [–.28, .38]
E20	–.01 (.03) [–.06, .05]	.19 (.08) [.07, .31]	.01 (.23) [–.35, .37]	1.56 (.23) [1.19, 1.92]
E21	–.01 (.03) [–.05, .04]	.33 (.07) [.22, .44]	1.33 (.21) [1.00, 1.67]	.14 (.21) [–.20, .48]
E22	.00 (.03) [–.04, .05]	.20 (.08) [.07, .31]	1.44 (.21) [1.11, 1.77]	.22 (.21) [–.11, .54]
E23	.00 (.03) [–.05, .05]	.06 (.09) [–.09, .20]	.57 (.22) [.22, .92]	1.04 (.21) [.70, 1.39]
E24	–.01 (.02) [–.05, .03]	.30 (.07) [.19, .41]	.79 (.21) [.45, 1.12]	.53 (.21) [.20, .87]
E25	–.01 (.03) [–.05, .04]	.30 (.07) [.18, .41]	1.01 (.21) [.68, 1.34]	–.03 (.22) [–.39, .32]

**Table E.3:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M2 for L1 Mandarin speakers at the individual level.

<b>Par.</b>	<b>intercept</b>	<b>high – (high + low)/2</b>	<b>asp – unasp</b>	<b>unasp – son</b>
M1	–.01 (.03) [–.05, .04]	.44 (.08) [.31, .57]	1.12 (.19) [.83, 1.42]	.41 (.12) [.24, .61]
M2	–.01 (.03) [–.06, .03]	.45 (.08) [.32, .57]	.50 (.19) [.19, .80]	.39 (.12) [.21, .58]
M3	–.01 (.03) [–.07, .04]	.44 (.09) [.30, .58]	.22 (.23) [–.16, .57]	.33 (.13) [.10, .53]
M4	–.01 (.03) [–.05, .03]	.54 (.08) [.42, .67]	.74 (.18) [.45, 1.03]	.32 (.12) [.12, .48]
M5	–.01 (.03) [–.05, .04]	.32 (.08) [.19, .45]	.88 (.19) [.59, 1.18]	.36 (.11) [.19, .54]
M6	–.01 (.03) [–.05, .03]	.54 (.08) [.42, .67]	.71 (.18) [.42, 1.00]	.34 (.11) [.15, .51]
M7	–.01 (.03) [–.05, .03]	.51 (.08) [.38, .63]	.78 (.18) [.50, 1.07]	.41 (.12) [.25, .62]
M8	–.01 (.03) [–.05, .03]	.55 (.09) [.41, .68]	.66 (.20) [.34, .99]	.33 (.12) [.11, .50]
M9	–.01 (.03) [–.05, .04]	.29 (.08) [.15, .42]	1.20 (.19) [.90, 1.51]	.34 (.12) [.15, .52]
M10	–.01 (.03) [–.05, .03]	.52 (.08) [.39, .65]	.89 (.18) [.61, 1.18]	.40 (.11) [.23, .58]
M11	–.01 (.03) [–.05, .03]	.52 (.08) [.40, .65]	.91 (.19) [.60, 1.21]	.42 (.12) [.26, .63]
M12	–.01 (.03) [–.05, .04]	.41 (.08) [.29, .54]	.93 (.19) [.62, 1.22]	.46 (.14) [.29, .72]
M13	–.01 (.03) [–.06, .03]	.55 (.08) [.42, .68]	.34 (.18) [.04, .62]	.38 (.11) [.20, .56]
M14	–.01 (.03)	.26 (.08)	.89 (.18)	.27 (.13)

**Table E.3 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], for parameter values estimated by M2 for L1 Mandarin speakers at the individual level.

<b>Par.</b>	<b>intercept</b>	<b>high – (high + low)/2</b>	<b>asp – unasp</b>	<b>unasp – son</b>
	[−.05, .04]	[.13, .39]	[.60, 1.17]	[.03, .44]
M15	−.01 (.03)	.60 (.09)	.62 (.20)	.44 (.13)
	[−.05, .03]	[.47, .74]	[.31, .94]	[.27, .69]
M16	−.01 (.03)	.24 (.08)	1.18 (.19)	.25 (.14)
	[−.05, .04]	[.10, .36]	[.87, 1.47]	[.01, .43]
M17	−.01 (.03)	.40 (.08)	1.09 (.18)	.36 (.11)
	[−.05, .04]	[.27, .52]	[.81, 1.37]	[.19, .54]
M18	−.01 (.03)	.40 (.08)	.75 (.19)	.34 (.11)
	[−.05, .03]	[.28, .52]	[.46, 1.04]	[.16, .49]
M19	−.01 (.03)	.55 (.08)	.34 (.19)	.41 (.12)
	[−.06, .03]	[.43, .68]	[.02, .64]	[.24, .62]
M20	−.01 (.03)	.54 (.08)	.73 (.19)	.32 (.12)
	[−.06, .03]	[.41, .68]	[.42, 1.03]	[.11, .49]
M21	.00 (.03)	.24 (.08)	1.39 (.19)	.30 (.13)
	[−.05, .04]	[.11, .37]	[1.09, 1.71]	[.08, .47]
M22	−.01 (.03)	.46 (.08)	.62 (.19)	.40 (.11)
	[−.05, .03]	[.34, .59]	[.31, .92]	[.25, .59]
M23	−.01 (.03)	.37 (.08)	1.05 (.18)	.32 (.11)
	[−.05, .03]	[.25, .50]	[.76, 1.34]	[.12, .48]
M24	−.01 (.03)	.30 (.08)	1.02 (.20)	.33 (.11)
	[−.05, .04]	[.16, .43]	[.71, 1.34]	[.15, .51]
M25	−.01 (.03)	.30 (.08)	.96 (.19)	.31 (.12)
	[−.05, .04]	[.17, .43]	[.66, 1.26]	[.11, .47]

**Table E.4:** Summary of posterior distributions for individual speakers' production post-stop F0 weights, in terms of mean (sd) [89% CrI].

<b>Par.</b>	<b>Post-stop F0 wt.</b>	<b>Par.</b>	<b>Post-stop F0 wt.</b>
E1	.88 (.26) [.46, 1.29]	M1	1.13 (.25) [.76, 1.55]
E2	1.28 (.27) [.86, 1.73]	M2	.59 (.24) [.17, .93]
E3	.99 (.30) [.51, 1.46]	M3	.55 (.24) [.14, .91]
E4	.64 (.25) [.24, 1.03]	M4	.90 (.22) [.55, 1.24]
E5	.69 (.26) [.28, 1.10]	M5	.83 (.22) [.47, 1.17]
E6	1.24 (.29) [.78, 1.70]	M6	.82 (.21) [.48, 1.15]
E7	.81 (.26) [.39, 1.24]	M7	.81 (.21) [.45, 1.14]
E8	.71 (.26) [.29, 1.13]	M8	.82 (.23) [.45, 1.19]
E9	1.34 (.27) [.92, 1.78]	M9	1.06 (.25) [.70, 1.49]
E10	.37 (.27) [−.06, .80]	M10	.87 (.21) [.54, 1.20]
E11	1.40 (.28) [.97, 1.85]	M11	.89 (.21) [.56, 1.24]
E12	1.38 (.29) [.93, 1.84]	M12	.87 (.23) [.48, 1.23]
E13	1.51 (.30) [1.04, 1.99]	M13	.55 (.24) [.13, .90]
E14	.69 (.26) [.25, 1.08]	M14	.89 (.21) [.57, 1.25]
E15	1.88 (.34) [1.38, 2.44]	M15	.74 (.23) [.35, 1.08]
E16	.42 (.27) [−.02, .83]	M16	1.17 (.26) [.79, 1.62]
E17	.70 (.27) [.28, 1.14]	M17	1.09 (.24) [.74, 1.49]
E18	.38 (.27) [−.07, .80]	M18	.82 (.22) [.45, 1.17]
E19	1.07 (.27) [.64, 1.51]	M19	.50 (.25) [.07, .86]
E20	.00 (.32) [−.52, .51]	M20	.94 (.24) [.58, 1.34]
E21	1.63 (.29) [1.19, 2.09]	M21	1.36 (.31) [.90, 1.89]
E22	1.68 (.30) [1.21, 2.17]	M22	.74 (.23) [.35, 1.08]
E23	.64 (.28) [.18, 1.09]	M23	1.05 (.24) [.70, 1.45]
E24	.82 (.26) [.41, 1.23]	M24	1.02 (.24) [.66, 1.42]
E25	1.17 (.28) [.72, 1.61]	M25	.90 (.22) [.57, 1.26]

**Table E.5:** Summary of posterior distributions for individual guessing probabilities, in terms of mean (sd) [89% CrI].

<b>Par.</b>	<b>Guessing prob.</b>	<b>Par.</b>	<b>Guessing prob.</b>
E1	.01 (.01) [.00, .02]	M1	.01 (.01) [.00, .02]
E2	.01 (.01) [.00, .02]	M2	.01 (.01) [.00, .02]
E3	.01 (.01) [.00, .03]	M3	.01 (.01) [.00, .02]

**Table E.5 continued:** Summary of posterior distributions for individual guessing probabilities, in terms of mean (sd) [89% CrI].

<b>Par.</b>	<b>Guessing prob.</b>	<b>Par.</b>	<b>Guessing prob.</b>
E4	.04 (.02) [.02, .06]	M4	.03 (.02) [.01, .07]
E5	.01 (.01) [.01, .02]	M5	.02 (.01) [.00, .04]
E6	.01 (.01) [.00, .03]	M6	.01 (.01) [.00, .03]
E7	.01 (.01) [.00, .02]	M7	.02 (.01) [.00, .04]
E8	.01 (.01) [.00, .02]	M8	.01 (.01) [.00, .04]
E9	.01 (.01) [.00, .02]	M9	.01 (.01) [.00, .02]
E10	.01 (.01) [.00, .03]	M10	.01 (.01) [.00, .02]
E11	.01 (.01) [.00, .02]	M11	.01 (.01) [.00, .02]
E12	.01 (.01) [.00, .02]	M12	.09 (.02) [.05, .13]
E13	.03 (.02) [.01, .06]	M13	.01 (.01) [.00, .02]
E14	.01 (.01) [.00, .02]	M14	.01 (.01) [.00, .02]
E15	.03 (.01) [.01, .05]	M15	.03 (.01) [.01, .06]
E16	.02 (.02) [.00, .05]	M16	.02 (.01) [.00, .03]
E17	.01 (.01) [.00, .03]	M17	.01 (.01) [.00, .03]
E18	.01 (.01) [.00, .02]	M18	.01 (.01) [.00, .03]
E19	.01 (.01) [.00, .03]	M19	.01 (.01) [.00, .03]
E20	.01 (.01) [.00, .02]	M20	.02 (.01) [.01, .04]
E21	.01 (.01) [.00, .02]	M21	.01 (.01) [.00, .02]
E22	.01 (.01) [.00, .02]	M22	.03 (.01) [.01, .06]
E23	.01 (.01) [.00, .02]	M23	.01 (.01) [.00, .03]
E24	.01 (.01) [.00, .02]	M24	.01 (.01) [.00, .02]
E25	.01 (.01) [.00, .02]	M25	.01 (.01) [.00, .03]

**Table E.6:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 English listeners.

<b>Par.</b>	<b>intercept</b>	<b>VOT wt.</b>	<b>F0 wt.</b>	<b>tone wt.</b>
E1	13.11 (1.67)	15.42 (1.82)	1.34 (.20)	.49 (.19)

**Table E.6 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 English listeners.

<b>Par.</b>	<b>intercept</b>	<b>VOT wt.</b>	<b>F0 wt.</b>	<b>tone wt.</b>
	[10.57, 15.87]	[12.63, 18.41]	[1.04, 1.64]	[.17, .79]
E2	12.69 (1.60)	17.24 (1.94)	1.36 (.22)	.50 (.22)
	[10.60, 15.58]	[14.67, 20.66]	[1.01, 1.69]	[.15, .85]
E3	12.89 (1.67)	16.90 (1.94)	1.36 (.21)	.61 (.24)
	[10.65, 15.89]	[14.32, 20.42]	[1.02, 1.68]	[.31, 1.04]
E4	12.10 (1.50)	17.00 (2.01)	1.36 (.22)	.47 (.24)
	[10.06, 14.62]	[14.29, 20.36]	[1.01, 1.70]	[.05, .82]
E5	12.81 (1.66)	17.48 (2.00)	1.38 (.21)	.58 (.23)
	[10.66, 15.78]	[14.83, 21.08]	[1.05, 1.71]	[.27, .99]
E6	12.14 (1.55)	18.04 (2.23)	1.36 (.23)	.55 (.24)
	[10.03, 14.90]	[15.07, 21.94]	[.97, 1.70]	[.21, .97]
E7	13.71 (1.82)	15.03 (1.94)	1.37 (.20)	.55 (.19)
	[10.96, 16.68]	[12.09, 18.16]	[1.06, 1.67]	[.27, .88]
E8	13.21 (1.65)	15.40 (1.81)	1.44 (.20)	.52 (.19)
	[10.76, 15.98]	[12.71, 18.32]	[1.15, 1.77]	[.24, .83]
E9	12.75 (1.66)	15.29 (1.86)	1.43 (.19)	.56 (.20)
	[10.20, 15.32]	[12.43, 18.13]	[1.15, 1.74]	[.27, .90]
E10	10.71 (1.28)	15.85 (1.70)	1.40 (.21)	.51 (.22)
	[8.74, 12.77]	[13.21, 18.47]	[1.07, 1.73]	[.17, .86]
E11	9.41 (1.26)	15.15 (2.02)	1.47 (.25)	.51 (.23)
	[7.36, 11.40]	[11.86, 18.28]	[1.13, 1.90]	[.15, .88]
E12	13.58 (1.81)	14.54 (1.90)	1.41 (.19)	.44 (.20)
	[10.82, 16.47]	[11.60, 17.57]	[1.12, 1.73]	[.09, .72]
E13	11.00 (2.49)	13.31 (2.87)	1.49 (.24)	.46 (.23)
	[6.61, 14.43]	[8.20, 17.15]	[1.18, 1.93]	[.09, .81]
E14	12.13 (1.44)	17.35 (1.94)	1.39 (.21)	.57 (.24)
	[10.08, 14.73]	[14.70, 20.87]	[1.07, 1.72]	[.24, .99]
E15	12.98 (1.85)	16.17 (2.16)	1.43 (.21)	.67 (.26)

**Table E.6 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 English listeners.

<b>Par.</b>	<b>intercept</b>	<b>VOT wt.</b>	<b>F0 wt.</b>	<b>tone wt.</b>
	[10.39, 16.22]	[12.97, 19.87]	[1.14, 1.76]	[.36, 1.16]
E16	8.25 (1.71)	14.92 (3.07)	1.39 (.26)	.31 (.28)
	[5.28, 10.86]	[9.69, 19.60]	[.99, 1.82]	[-.19, .69]
E17	12.38 (2.16)	12.31 (2.26)	1.43 (.22)	.36 (.24)
	[9.07, 15.97]	[8.90, 16.11]	[1.12, 1.79]	[-.08, .68]
E18	13.95 (1.84)	14.89 (1.94)	1.39 (.20)	.59 (.20)
	[11.17, 16.95]	[11.93, 18.07]	[1.09, 1.70]	[.30, .95]
E19	12.06 (1.46)	16.80 (1.86)	1.40 (.21)	.59 (.23)
	[10.01, 14.56]	[14.20, 20.01]	[1.09, 1.74]	[.28, .99]
E20	11.42 (1.26)	16.79 (1.76)	1.40 (.22)	.45 (.22)
	[9.59, 13.51]	[14.26, 19.78]	[1.08, 1.75]	[.08, .76]
E21	12.83 (1.67)	17.48 (2.01)	1.37 (.21)	.58 (.23)
	[10.63, 15.80]	[14.86, 21.05]	[1.03, 1.70]	[.26, .98]
E22	14.31 (1.87)	14.62 (1.94)	1.43 (.21)	.55 (.20)
	[11.46, 17.33]	[11.68, 17.83]	[1.13, 1.80]	[.26, .89]
E23	13.23 (1.70)	16.31 (1.91)	1.40 (.20)	.64 (.24)
	[10.84, 16.15]	[13.62, 19.70]	[1.10, 1.72]	[.34, 1.09]
E24	10.61 (1.21)	17.24 (2.07)	1.35 (.25)	.48 (.23)
	[8.81, 12.62]	[14.20, 20.75]	[.95, 1.70]	[.10, .83]
E25	12.25 (1.47)	16.66 (1.75)	1.38 (.21)	.60 (.23)
	[10.19, 14.71]	[14.14, 19.62]	[1.07, 1.71]	[.31, 1.01]

**Table E.7:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 Mandarin listeners.

<b>Par.</b>	<b>intercept</b>	<b>VOT wt.</b>	<b>F0 wt.</b>	<b>tone wt.</b>
M1	10.54 (1.73)	14.75 (2.27)	1.46 (.48)	.57 (.24)

**Table E.7 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 Mandarin listeners.

<b>Par.</b>	<b>intercept</b>	<b>VOT wt.</b>	<b>F0 wt.</b>	<b>tone wt.</b>
	[8.04, 13.59]	[11.44, 18.78]	[.84, 2.33]	[.22, .98]
M2	12.45 (2.13)	17.88 (2.76)	1.02 (.44)	.43 (.26)
	[9.54, 16.13]	[13.91, 22.62]	[.35, 1.73]	[−.01, .83]
M3	10.79 (1.58)	17.72 (2.79)	.91 (.39)	.37 (.25)
	[8.49, 13.48]	[13.75, 22.54]	[.29, 1.51]	[−.07, .71]
M4	6.69 (2.07)	13.51 (4.38)	1.06 (.36)	.38 (.25)
	[3.79, 10.23]	[7.60, 21.23]	[.55, 1.67]	[−.06, .72]
M5	10.87 (1.71)	16.69 (2.62)	.92 (.38)	.39 (.26)
	[8.43, 13.73]	[12.89, 21.20]	[.32, 1.53]	[−.06, .72]
M6	10.14 (1.50)	15.89 (2.40)	.91 (.36)	.43 (.24)
	[7.95, 12.66]	[12.38, 19.97]	[.33, 1.48]	[.03, .79]
M7	9.95 (2.18)	13.02 (2.60)	1.08 (.35)	.44 (.23)
	[7.00, 13.76]	[9.39, 17.45]	[.56, 1.66]	[.04, .76]
M8	8.70 (1.59)	16.37 (3.28)	.53 (.35)	.30 (.26)
	[6.23, 11.32]	[11.41, 21.96]	[−.06, 1.05]	[−.16, .63]
M9	10.32 (1.62)	14.80 (2.18)	1.26 (.41)	.57 (.24)
	[7.93, 13.12]	[11.64, 18.49]	[.72, 1.99]	[.24, .98]
M10	9.06 (1.49)	16.60 (2.96)	1.20 (.39)	.67 (.29)
	[6.85, 11.57]	[12.30, 21.71]	[.68, 1.90]	[.31, 1.21]
M11	11.41 (1.74)	16.96 (2.46)	1.05 (.42)	.51 (.24)
	[8.93, 14.38]	[13.43, 21.17]	[.44, 1.75]	[.16, .91]
M12	8.83 (1.54)	16.43 (3.02)	.66 (.37)	.43 (.24)
	[6.48, 11.34]	[11.90, 21.36]	[.04, 1.19]	[.06, .78]
M13	12.46 (2.00)	16.77 (2.43)	1.12 (.38)	.46 (.24)
	[9.65, 15.90]	[13.28, 20.91]	[.55, 1.76]	[.04, .83]
M14	12.90 (2.04)	17.74 (2.54)	1.20 (.43)	.34 (.29)
	[10.06, 16.54]	[14.10, 22.19]	[.60, 1.95]	[−.18, .69]
M15	11.34 (1.84)	17.68 (2.94)	.86 (.41)	.38 (.26)

**Table E.7 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 Mandarin listeners.

<b>Par.</b>	<b>intercept</b>	<b>VOT wt.</b>	<b>F0 wt.</b>	<b>tone wt.</b>
	[8.69, 14.45]	[13.40, 22.69]	[.16, 1.48]	[−.07, .74]
M16	11.38 (1.79)	16.91 (2.52)	.92 (.39)	.49 (.25)
	[8.87, 14.51]	[13.26, 21.22]	[.25, 1.51]	[.12, .88]
M17	11.60 (1.76)	17.24 (2.53)	1.03 (.39)	.38 (.26)
	[9.10, 14.62]	[13.61, 21.50]	[.43, 1.66]	[−.09, .73]
M18	12.81 (2.02)	18.19 (2.69)	1.05 (.43)	.44 (.26)
	[10.03, 16.36]	[14.53, 22.90]	[.40, 1.74]	[.01, .84]
M19	8.19 (1.49)	15.99 (3.14)	.75 (.31)	.52 (.22)
	[5.91, 10.59]	[11.27, 21.16]	[.25, 1.23]	[.20, .88]
M20	12.33 (2.20)	16.53 (2.81)	1.22 (.41)	.55 (.25)
	[9.16, 16.14]	[12.26, 21.22]	[.65, 1.93]	[.19, .97]
M21	10.11 (1.47)	15.18 (2.16)	1.29 (.41)	.51 (.23)
	[7.89, 12.58]	[11.99, 18.85]	[.76, 2.04]	[.17, .88]
M22	9.73 (1.51)	16.85 (2.85)	.57 (.41)	.44 (.24)
	[7.46, 12.28]	[12.68, 21.63]	[−.14, 1.14]	[.06, .80]
M23	12.17 (2.18)	14.16 (2.45)	1.38 (.39)	.61 (.24)
	[9.02, 15.89]	[10.60, 18.32]	[.83, 2.08]	[.27, 1.03]
M24	12.45 (1.84)	18.99 (2.81)	1.03 (.43)	.35 (.29)
	[9.88, 15.59]	[15.01, 23.84]	[.39, 1.76]	[−.18, .72]
M25	6.15 (1.15)	6.87 (1.24)	.85 (.31)	.69 (.27)
	[4.55, 8.11]	[5.14, 8.97]	[.37, 1.34]	[.33, 1.16]

## Appendix F

# Supporting Materials for Chapter 5

### F.1 Participant Demographic Information

**Table F.1:** Demographic information of the L1 English participants in the CL condition. N = non-binary; Am. = American English; Br. = British English; Ca. = Canadian English.

ID	Gender	Age	Eng. variety	Other lang.
1	F	23	Ca.	French
2	M	22	Am./Ca./Jamaican	Patois
3	F	22	Ca.	Spanish/French
4	M	22	Am./Ca.	Spanish
5	F	21	Am.	Korean
6	M	25	Am./Western New England	Spanish
7	F	20	Ca.	Swedish
8	F	19	Ca.	Trinidad Creole/French/Spanish
9	F	19	Am.	French/Spanish/Greek
10	F	19	Am./Br./Ca.	Japanese
11	M	20	Ca.	French

**Table F.1 continued:** Demographic information of the L1 English participants in the CL condition. N = non-binary; Am. = American English; Br. = British English; Ca. = Canadian English.

<b>ID</b>	<b>Gender</b>	<b>Age</b>	<b>Eng. variety</b>	<b>Other lang.</b>
12	F	18	Am./Br./Ca.	
13	F	18	Ca.	French/Italian
14	F	18	Ca.	Spanish
15	F	17	Ca.	
16	F	23	Ca.	
17	F	21	Am.	Spanish
18	N	20	Am./Ca.	Korean
19	N	19	Ca.	French
20	F	33	Ca.	Spanish
21	F	17	Ca.	
22	F	17	Ca.	
23	F	20	Am./Ca.	
24	M	23	Br./Ca.	
25	F	19	Ca.	

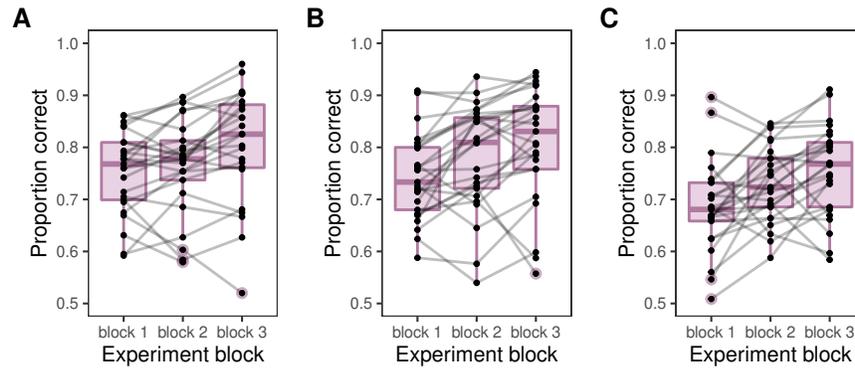
**Table F.2:** Demographic information of the L1 Mandarin-L2 English participants in the CL condition. BLP = Bilingual Language Profile Score (see Section 6.4.2); Am. = American English; Br. = British English; Ca. = Canadian English.

<b>ID</b>	<b>Gender</b>	<b>Age</b>	<b>BLP</b>	<b>Grow-up place</b>	<b>Eng. variety</b>	<b>Other lang.</b>
1	F	26	81.1	Fuzhou	Am./Ca.	
2	F	24	44.8	Xi'an	Ca.	
3	F	22	59.8	Shenyang	Ca.	
4	F	21	9.0	Shanghai	Am./Ca.	
5	F	21	80.7	Shenzhen	Am./Ca.	
6	F	21	56.8	Shanghai	Am.	
7	F	20	70.9	Nanchang	Ca.	French

**Table F.2 continued:** Demographic information of the L1 Mandarin-L2 English participants in the CL condition. BLP = Bilingual Language Profile Score (see Section 6.4.2); Am. = American English; Br. = British English; Ca. = Canadian English.

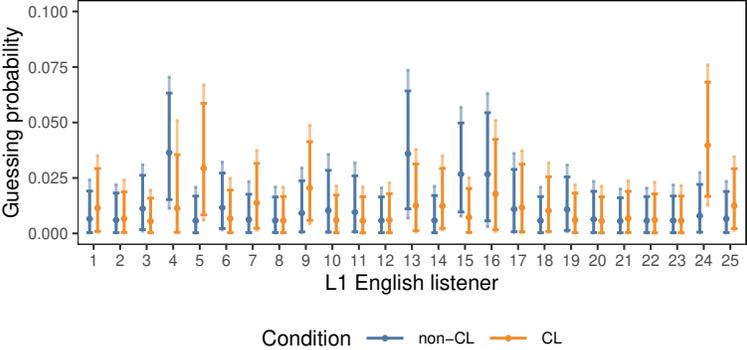
<b>ID</b>	<b>Gender</b>	<b>Age</b>	<b>BLP</b>	<b>Grow-up place</b>	<b>Eng. variety</b>	<b>Other lang.</b>
8	F	20	55.7	Tianjin	Am.	
9	M	20	40.2	Nanjing	Am./Br./Ca.	
10	F	20	96.0	Zhejiang	Ca.	
11	F	20	112.9	Tianjin	Am./Ca.	
12	F	20	74.4	Zhangjiagang	Am./Br./Ca.	Korean
13	M	20	91.2	Shanghai	Am.	Taizhouhua
14	F	19	96.7	Beijing	Am.	
15	F	20	122.1	Wuxi	Chinglish	
16	F	19	42.3	Beijing	Am./Ca.	
17	F	19	58.5	Shenyang	Br./Ca.	Sichuanese
18	M	20	34.5	Kunming	Am.	French
19	F	19	121.6	Shijiazhuang	Am.	
20	F	18	82.9	Beijing	Am.	
21	F	18	80.5	Shanghai	Am./Ca./Newfoundland	
22	F	18	48.7	Shanghai	Ca.	
23	F	18	89.9	Shanghai	Br./Ca.	
24	M	19	98.4	[China]	Am./Ca.	
25	F	18	97.1	Wuhan	Am.	Korean

## F.2 Accuracy of the Visual Search Task



**Figure F.1:** Accuracy in the visual search task, based on both “correct” target and filler trials (see Section 5.2.5 for detail). A score was estimated for each listener in each block. The scores from the same listeners are connected by a line. **A.** Results for L1 English listeners who also participated in the non-CL version. **B.** Results for L1 English listeners who only did the CL version. **C.** Results for L1 Mandarin listeners.

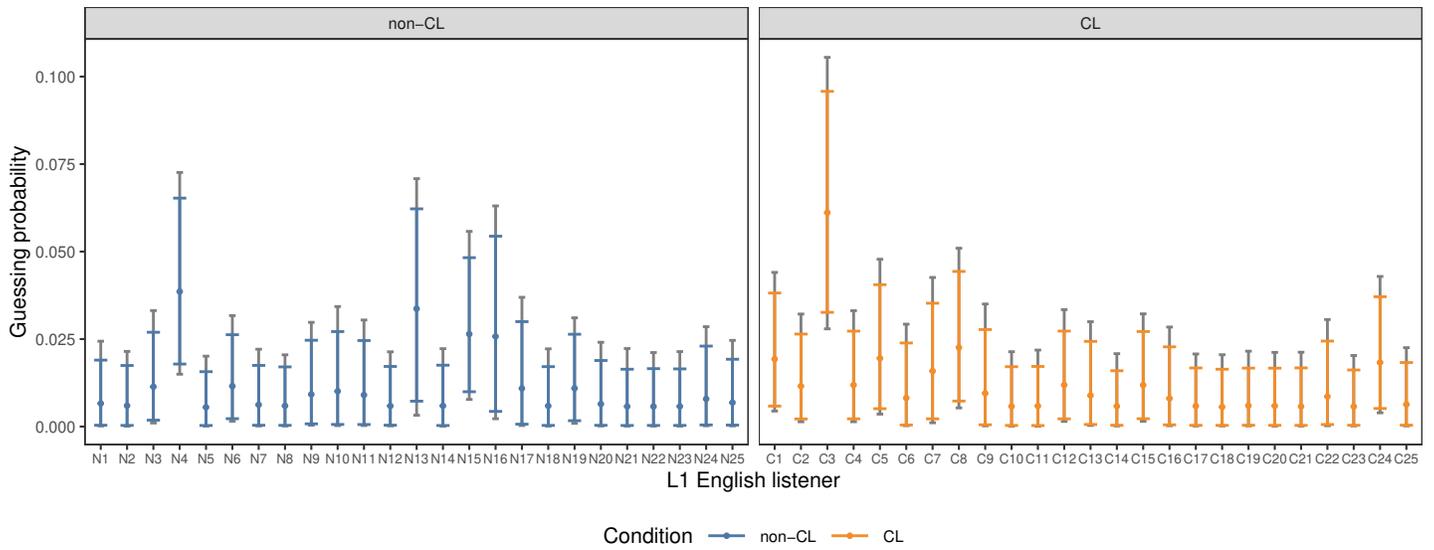
### F.3 Individual Guessing Probabilities



**Figure F.2:** L1 English listeners’ guessing probabilities in each experimental condition. The dots mark the posterior means. The 89% CrIs are marked by the inner error bars, and the 95% CrIs are marked by the outer error bars.

**Table F.3:** Summary of posterior distributions for guessing probabilities, in terms of mean (sd) [89% CrI].

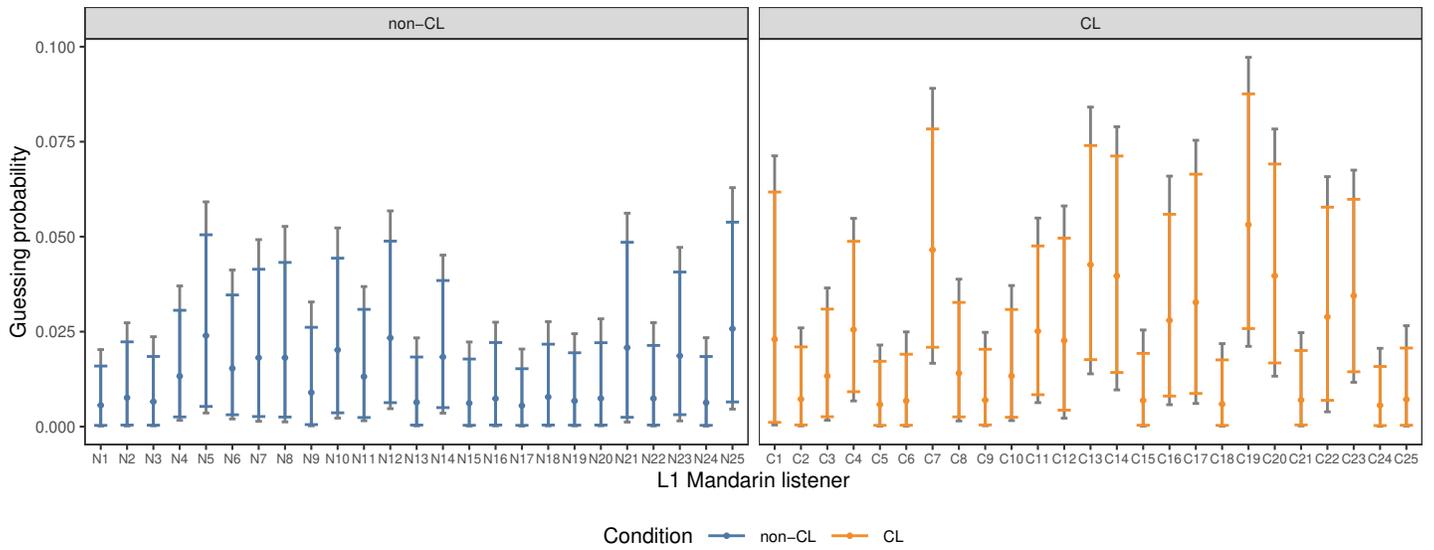
<b>Par.</b>	<b>Guessing prob. (non-CL)</b>	<b>Guessing prob. (CL)</b>
E1	.01 (.01) [.00, .02]	.01 (.01) [.00, .03]
E2	.01 (.01) [.00, .02]	.01 (.01) [.00, .02]
E3	.01 (.01) [.00, .03]	.01 (.01) [.00, .02]
E4	.04 (.02) [.02, .06]	.01 (.01) [.00, .04]
E5	.01 (.01) [.00, .02]	.03 (.02) [.01, .06]
E6	.01 (.01) [.00, .03]	.01 (.01) [.00, .02]
E7	.01 (.01) [.00, .02]	.01 (.01) [.00, .03]
E8	.01 (.01) [.00, .02]	.01 (.01) [.00, .02]
E9	.01 (.01) [.00, .02]	.02 (.01) [.01, .04]
E10	.01 (.01) [.00, .03]	.01 (.01) [.00, .02]
E11	.01 (.01) [.00, .03]	.01 (.01) [.00, .02]
E12	.01 (.01) [.00, .02]	.01 (.01) [.00, .02]
E13	.04 (.02) [.01, .06]	.01 (.01) [.00, .03]
E14	.01 (.01) [.00, .02]	.01 (.01) [.00, .03]
E15	.03 (.01) [.01, .05]	.01 (.01) [.00, .02]
E16	.03 (.02) [.01, .05]	.02 (.01) [.00, .04]
E17	.01 (.01) [.00, .03]	.01 (.01) [.00, .03]
E18	.01 (.01) [.00, .02]	.01 (.01) [.00, .03]
E19	.01 (.01) [.00, .03]	.01 (.01) [.00, .02]
E20	.01 (.01) [.00, .02]	.01 (.01) [.00, .02]
E21	.01 (.01) [.00, .02]	.01 (.01) [.00, .02]
E22	.01 (.01) [.00, .02]	.01 (.01) [.00, .02]
E23	.01 (.01) [.00, .02]	.01 (.01) [.00, .02]
E24	.01 (.01) [.00, .02]	.04 (.02) [.02, .07]
E25	.01 (.01) [.00, .02]	.01 (.01) [.00, .03]



**Figure F.3:** Guessing probabilities for L1 English listeners from both groups. The posterior means are marked by the dots. The 89% CrIs are marked by the inner error bars, and the 95% CrIs are marked by the outer error bars.

**Table F.4:** Summary of posterior distributions for guessing probabilities, in terms of mean (sd) [89% CrI].

<b>Par.</b>	<b>Guessing prob. (non-CL)</b>	<b>Par.</b>	<b>Guessing prob. (CL)</b>
EN1	.01 (.01) [.00, .02]	EC1	.02 (.01) [.01, .04]
EN2	.01 (.01) [.00, .02]	EC2	.01 (.01) [.00, .03]
EN3	.01 (.01) [.00, .03]	EC3	.06 (.02) [.03, .10]
EN4	.04 (.02) [.02, .07]	EC4	.01 (.01) [.00, .03]
EN5	.01 (.01) [.00, .02]	EC5	.02 (.01) [.01, .04]
EN6	.01 (.01) [.00, .03]	EC6	.01 (.01) [.00, .02]
EN7	.01 (.01) [.00, .02]	EC7	.02 (.01) [.00, .04]
EN8	.01 (.01) [.00, .02]	EC8	.02 (.01) [.01, .04]
EN9	.01 (.01) [.00, .02]	EC9	.01 (.01) [.00, .03]
EN10	.01 (.01) [.00, .03]	EC10	.01 (.01) [.00, .02]
EN11	.01 (.01) [.00, .02]	EC11	.01 (.01) [.00, .02]
EN12	.01 (.01) [.00, .02]	EC12	.01 (.01) [.00, .03]
EN13	.03 (.02) [.01, .06]	EC13	.01 (.01) [.00, .02]
EN14	.01 (.01) [.00, .02]	EC14	.01 (.01) [.00, .02]
EN15	.03 (.01) [.01, .05]	EC15	.01 (.01) [.00, .03]
EN16	.03 (.02) [.00, .05]	EC16	.01 (.01) [.00, .02]
EN17	.01 (.01) [.00, .03]	EC17	.01 (.01) [.00, .02]
EN18	.01 (.01) [.00, .02]	EC18	.01 (.01) [.00, .02]
EN19	.01 (.01) [.00, .03]	EC19	.01 (.01) [.00, .02]
EN20	.01 (.01) [.00, .02]	EC20	.01 (.01) [.00, .02]
EN21	.01 (.01) [.00, .02]	EC21	.01 (.01) [.00, .02]
EN22	.01 (.01) [.00, .02]	EC22	.01 (.01) [.00, .02]
EN23	.01 (.01) [.00, .02]	EC23	.01 (.01) [.00, .02]
EN24	.01 (.01) [.00, .02]	EC24	.02 (.01) [.01, .04]
EN25	.01 (.01) [.00, .02]	EC25	.01 (.01) [.00, .02]



**Figure F.4:** Guessing probabilities for L1 Mandarin listeners from both groups. The posterior means are marked by the dots. The 89% CrIs are marked by the inner error bars, and the 95% CrIs are marked by the outer error bars.

**Table F.5:** Summary of posterior distributions for guessing probabilities, in terms of mean (sd) [89% CrI].

<b>Par.</b>	<b>Guessing prob. (non-CL)</b>	<b>Par.</b>	<b>Guessing prob. (CL)</b>
MN1	.01 (.01) [.00, .02]	MC1	.02 (.02) [.00, .06]
MN2	.01 (.01) [.00, .02]	MC2	.01 (.01) [.00, .02]
MN3	.01 (.01) [.00, .02]	MC3	.01 (.01) [.00, .03]
MN4	.01 (.01) [.00, .03]	MC4	.03 (.01) [.01, .05]
MN5	.02 (.01) [.01, .05]	MC5	.01 (.01) [.00, .02]
MN6	.02 (.01) [.00, .03]	MC6	.01 (.01) [.00, .02]
MN7	.02 (.01) [.00, .04]	MC7	.05 (.02) [.02, .08]
MN8	.02 (.01) [.00, .04]	MC8	.01 (.01) [.00, .03]
MN9	.01 (.01) [.00, .03]	MC9	.01 (.01) [.00, .02]
MN10	.02 (.01) [.00, .04]	MC10	.01 (.01) [.00, .03]
MN11	.01 (.01) [.00, .03]	MC11	.03 (.01) [.01, .05]
MN12	.02 (.01) [.01, .05]	MC12	.02 (.01) [.00, .05]
MN13	.01 (.01) [.00, .02]	MC13	.04 (.02) [.02, .07]
MN14	.02 (.01) [.01, .04]	MC14	.04 (.02) [.01, .07]
MN15	.01 (.01) [.00, .02]	MC15	.01 (.01) [.00, .02]
MN16	.01 (.01) [.00, .02]	MC16	.03 (.02) [.01, .06]
MN17	.01 (.01) [.00, .02]	MC17	.03 (.02) [.01, .07]
MN18	.01 (.01) [.00, .02]	MC18	.01 (.01) [.00, .02]
MN19	.01 (.01) [.00, .02]	MC19	.05 (.02) [.03, .09]
MN20	.01 (.01) [.00, .02]	MC20	.04 (.02) [.02, .07]
MN21	.02 (.01) [.00, .05]	MC21	.01 (.01) [.00, .02]
MN22	.01 (.01) [.00, .02]	MC22	.03 (.02) [.01, .06]
MN23	.02 (.01) [.00, .04]	MC23	.03 (.01) [.01, .06]
MN24	.01 (.01) [.00, .02]	MC24	.01 (.01) [.00, .02]
MN25	.03 (.02) [.01, .05]	MC25	.01 (.01) [.00, .02]

## F.4 Posterior Summaries of Individual-Level Parameters

**Table F.6:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 English listeners.

Par.	intercept (non-CL)	intercept (CL)	VOT wt. (non-CL)	VOT wt. (CL)	F0 wt. (non-CL)	F0 wt. (CL)	tone wt. (non-CL)	tone wt. (CL)
E1	13.56 (1.69) [11.00, 16.31]	11.79 (2.08) [8.88, 15.49]	15.80 (1.82) [12.95, 18.70]	12.00 (2.21) [8.88, 15.49]	1.39 (.23) [1.00, 1.73]	1.33 (.35) [8.88, 15.49]	.40 (.24) [-.03, .74]	-.05 (.28) [8.88, 15.49]
E2	12.82 (1.41) [10.79, 15.19]	11.46 (1.70) [8.93, 14.36]	17.32 (1.66) [14.92, 20.11]	14.16 (1.99) [8.93, 14.36]	1.44 (.22) [1.08, 1.77]	1.47 (.36) [8.93, 14.36]	.46 (.25) [.05, .81]	.42 (.28) [8.93, 14.36]
E3	13.37 (1.53) [11.26, 16.00]	13.72 (2.09) [10.78, 17.32]	17.63 (1.77) [15.16, 20.75]	18.86 (2.69) [10.78, 17.32]	1.44 (.23) [1.07, 1.79]	1.28 (.43) [10.78, 17.32]	.77 (.29) [.38, 1.31]	.55 (.38) [10.78, 17.32]
E4	12.16 (1.78) [9.46, 14.99]	6.84 (1.82) [4.98, 9.55]	16.66 (2.28) [13.06, 20.04]	6.93 (2.02) [4.98, 9.55]	1.37 (.31) [.83, 1.79]	.80 (.36) [4.98, 9.55]	.18 (.38) [-.50, .69]	-.23 (.27) [4.98, 9.55]
E5	12.77 (1.50) [10.61, 15.35]	9.38 (1.85) [6.48, 12.39]	17.69 (1.86) [15.00, 20.89]	13.62 (2.77) [6.48, 12.39]	1.45 (.25) [1.06, 1.83]	1.80 (.52) [6.48, 12.39]	.57 (.29) [.12, 1.04]	.42 (.35) [6.48, 12.39]
E6	12.39 (1.36) [10.48, 14.74]	11.58 (1.63) [9.22, 14.33]	18.71 (2.03) [15.94, 22.25]	17.85 (2.57) [9.22, 14.33]	1.44 (.25) [1.03, 1.81]	1.34 (.43) [9.22, 14.33]	.65 (.29) [.23, 1.15]	.56 (.37) [9.22, 14.33]
E7	14.24 (1.73) [11.57, 17.10]	12.98 (2.09) [9.93, 16.61]	15.56 (1.85) [12.68, 18.54]	13.72 (2.27) [9.93, 16.61]	1.44 (.21) [1.09, 1.76]	1.68 (.37) [9.93, 16.61]	.56 (.21) [.23, .90]	.22 (.27) [9.93, 16.61]
E8	13.72 (1.61) [11.29, 16.35]	13.41 (1.97) [10.62, 16.75]	15.97 (1.73) [13.33, 18.80]	15.49 (2.21) [10.62, 16.75]	1.53 (.21) [1.22, 1.88]	1.66 (.37) [10.62, 16.75]	.55 (.21) [.22, .89]	.34 (.27) [10.62, 16.75]
E9	13.41 (1.61) [10.85, 16.04]	12.46 (1.98) [9.56, 15.75]	16.09 (1.80) [13.25, 18.95]	15.25 (2.29) [9.56, 15.75]	1.51 (.21) [1.19, 1.85]	1.51 (.37) [9.56, 15.75]	.60 (.22) [.27, .98]	.20 (.28) [9.56, 15.75]
E10	11.41 (1.27) [9.47, 13.47]	12.40 (1.79) [9.80, 15.49]	16.71 (1.59) [14.28, 19.25]	16.48 (2.28) [9.80, 15.49]	1.51 (.23) [1.18, 1.87]	1.68 (.42) [9.80, 15.49]	.59 (.26) [.20, 1.01]	.67 (.33) [9.80, 15.49]
E11	10.20 (1.24) [8.18, 12.20]	13.22 (2.15) [10.17, 16.89]	16.15 (1.87) [13.03, 19.04]	17.33 (2.69) [10.17, 16.89]	1.61 (.27) [1.25, 2.11]	1.59 (.43) [10.17, 16.89]	.63 (.29) [.23, 1.15]	.81 (.40) [10.17, 16.89]

**Table F.6 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 English listeners.

Par.	intercept (non-CL)	intercept (CL)	VOT wt. (non-CL)	VOT wt. (CL)	F0 wt. (non-CL)	F0 wt. (CL)	tone wt. (non-CL)	tone wt. (CL)
E12	14.41 (1.78) [11.71, 17.36]	13.43 (2.11) [10.41, 17.14]	15.39 (1.88) [12.51, 18.41]	14.01 (2.25) [10.41, 17.14]	1.48 (.21) [1.16, 1.80]	1.35 (.32) [10.41, 17.14]	.43 (.22) [.06, .76]	.19 (.26) [10.41, 17.14]
E13	12.39 (2.26) [8.24, 15.63]	11.70 (2.11) [8.72, 15.33]	14.86 (2.55) [10.02, 18.38]	13.40 (2.41) [8.72, 15.33]	1.58 (.25) [1.24, 2.02]	1.56 (.37) [8.72, 15.33]	.48 (.24) [.07, .84]	-.05 (.28) [8.72, 15.33]
E14	12.56 (1.36) [10.63, 14.88]	12.96 (1.94) [10.19, 16.41]	17.94 (1.77) [15.44, 20.93]	17.73 (2.49) [10.19, 16.41]	1.47 (.24) [1.10, 1.84]	1.02 (.47) [10.19, 16.41]	.67 (.28) [.28, 1.14]	.57 (.37) [10.19, 16.41]
E15	13.55 (1.62) [11.15, 16.15]	12.70 (1.94) [9.92, 15.97]	16.99 (1.87) [14.18, 20.01]	16.89 (2.38) [9.92, 15.97]	1.52 (.23) [1.18, 1.90]	1.28 (.40) [9.92, 15.97]	.79 (.31) [.41, 1.35]	.34 (.34) [9.92, 15.97]
E16	9.15 (1.47) [6.64, 11.39]	8.26 (1.79) [5.76, 11.32]	16.66 (2.63) [12.09, 20.67]	12.76 (2.67) [5.76, 11.32]	1.48 (.28) [1.07, 1.92]	1.59 (.49) [5.76, 11.32]	.28 (.32) [-.28, .71]	.48 (.37) [5.76, 11.32]
E17	13.64 (2.27) [10.18, 17.41]	12.41 (2.50) [8.97, 16.88]	13.63 (2.41) [10.00, 17.71]	12.60 (2.73) [8.97, 16.88]	1.50 (.23) [1.16, 1.87]	1.59 (.38) [8.97, 16.88]	.34 (.27) [-.14, .70]	-.04 (.30) [8.97, 16.88]
E18	14.73 (1.83) [11.86, 17.65]	12.95 (2.22) [9.59, 16.69]	15.78 (1.94) [12.74, 18.84]	15.04 (2.49) [9.59, 16.69]	1.44 (.22) [1.08, 1.77]	1.03 (.36) [9.59, 16.69]	.65 (.23) [.32, 1.06]	.06 (.28) [9.59, 16.69]
E19	12.45 (1.32) [10.50, 14.71]	13.03 (1.96) [10.24, 16.41]	17.11 (1.64) [14.67, 19.89]	16.05 (2.29) [10.24, 16.41]	1.50 (.22) [1.16, 1.84]	1.48 (.36) [10.24, 16.41]	.62 (.26) [.25, 1.07]	.64 (.32) [10.24, 16.41]
E20	12.02 (1.23) [10.18, 14.01]	13.14 (1.99) [10.37, 16.63]	17.75 (1.70) [15.26, 20.61]	18.81 (2.74) [10.37, 16.63]	1.50 (.25) [1.13, 1.89]	1.12 (.47) [10.37, 16.63]	.55 (.28) [.10, .97]	.50 (.38) [10.37, 16.63]
E21	12.98 (1.45) [11.01, 15.42]	13.03 (2.00) [10.15, 16.47]	17.74 (1.76) [15.27, 20.73]	16.94 (2.44) [10.15, 16.47]	1.47 (.23) [1.12, 1.83]	1.65 (.42) [10.15, 16.47]	.65 (.27) [.28, 1.13]	.84 (.38) [10.15, 16.47]
E22	14.96 (1.90) [12.03, 18.06]	14.36 (2.18) [11.24, 18.15]	15.25 (1.96) [12.21, 18.41]	15.00 (2.30) [11.24, 18.15]	1.52 (.22) [1.20, 1.88]	1.74 (.39) [11.24, 18.15]	.60 (.22) [.26, .97]	.29 (.28) [11.24, 18.15]

**Table F.6 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 English listeners.

Par.	intercept (non-CL)	intercept (CL)	VOT wt. (non-CL)	VOT wt. (CL)	F0 wt. (non-CL)	F0 wt. (CL)	tone wt. (non-CL)	tone wt. (CL)
E23	13.74 (1.64) [11.44, 16.53]	14.39 (2.23) [11.18, 18.15]	16.97 (1.81) [14.37, 19.93]	17.77 (2.61) [11.18, 18.15]	1.50 (.23) [1.17, 1.87]	1.59 (.40) [11.18, 18.15]	.77 (.29) [.40, 1.29]	.86 (.38) [11.18, 18.15]
E24	11.02 (1.16) [9.32, 12.93]	11.10 (2.03) [8.10, 14.57]	17.89 (1.88) [15.14, 21.07]	16.25 (2.97) [8.10, 14.57]	1.42 (.26) [.98, 1.79]	.99 (.46) [8.10, 14.57]	.50 (.27) [.05, .90]	.39 (.40) [8.10, 14.57]
E25	12.82 (1.45) [10.76, 15.31]	13.04 (1.89) [10.36, 16.31]	17.66 (1.74) [15.12, 20.59]	18.84 (2.63) [10.36, 16.31]	1.47 (.24) [1.10, 1.84]	1.17 (.47) [10.36, 16.31]	.78 (.32) [.38, 1.35]	.54 (.38) [10.36, 16.31]

**Table F.7:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 English listeners in the non-CL condition.

<b>Par.</b>	<b>intercept</b>	<b>VOT wt.</b>	<b>F0 wt.</b>	<b>tone wt.</b>
EN1	13.94 (1.78) [11.22, 16.83]	16.34 (1.92) [13.41, 19.36]	1.43 (.21) [1.07, 1.73]	.51 (.20) [.18, .82]
EN2	13.26 (1.62) [11.15, 16.17]	17.96 (1.93) [15.45, 21.45]	1.47 (.22) [1.12, 1.80]	.53 (.23) [.17, .87]
EN3	13.63 (1.76) [11.30, 16.72]	17.80 (2.01) [15.10, 21.34]	1.47 (.21) [1.12, 1.79]	.65 (.25) [.33, 1.11]
EN4	12.70 (1.54) [10.57, 15.31]	17.78 (2.01) [15.03, 21.29]	1.47 (.22) [1.13, 1.80]	.50 (.25) [.08, .85]
EN5	13.49 (1.73) [11.28, 16.67]	18.30 (2.09) [15.66, 22.13]	1.49 (.22) [1.16, 1.84]	.62 (.24) [.29, 1.05]
EN6	12.63 (1.52) [10.57, 15.33]	18.71 (2.21) [15.90, 22.70]	1.48 (.23) [1.12, 1.83]	.59 (.26) [.22, 1.05]
EN7	14.67 (1.80) [11.94, 17.60]	16.06 (1.91) [13.14, 19.17]	1.45 (.20) [1.11, 1.76]	.58 (.20) [.29, .92]
EN8	14.06 (1.69) [11.48, 16.85]	16.34 (1.82) [13.45, 19.30]	1.54 (.20) [1.25, 1.89]	.55 (.19) [.25, .86]
EN9	13.62 (1.74) [10.95, 16.37]	16.30 (1.94) [13.23, 19.33]	1.53 (.21) [1.24, 1.88]	.59 (.20) [.30, .92]
EN10	11.31 (1.37) [9.15, 13.44]	16.66 (1.79) [13.74, 19.42]	1.52 (.21) [1.19, 1.86]	.55 (.23) [.21, .94]
EN11	9.89 (1.30) [7.75, 11.93]	15.91 (2.05) [12.39, 18.98]	1.61 (.26) [1.26, 2.07]	.54 (.24) [.18, .93]
EN12	14.56 (1.88) [11.50, 17.43]	15.57 (1.99) [12.37, 18.54]	1.50 (.20) [1.19, 1.82]	.46 (.21) [.09, .74]
EN13	12.11 (2.60) [7.16, 15.60]	14.59 (2.97) [8.76, 18.38]	1.61 (.25) [1.29, 2.05]	.50 (.23) [.11, .84]
EN14	12.71 (1.47)	18.12 (1.94)	1.51 (.22)	.61 (.25)

**Table F.7 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 English listeners in the non-CL condition.

<b>Par.</b>	<b>intercept</b>	<b>VOT wt.</b>	<b>F0 wt.</b>	<b>tone wt.</b>
	[10.69, 15.34]	[15.57, 21.62]	[1.18, 1.85]	[.28, 1.06]
EN15	13.85 (1.84)	17.21 (2.13)	1.54 (.21)	.71 (.27)
	[11.26, 17.07]	[14.24, 20.92]	[1.24, 1.90]	[.38, 1.24]
EN16	8.76 (1.74)	15.78 (3.10)	1.52 (.27)	.35 (.29)
	[5.69, 11.28]	[10.29, 20.23]	[1.13, 1.96]	[−.16, .74]
EN17	13.38 (2.47)	13.34 (2.60)	1.52 (.22)	.37 (.26)
	[9.47, 17.30]	[9.30, 17.46]	[1.19, 1.88]	[−.10, .71]
EN18	14.97 (1.84)	15.98 (1.95)	1.48 (.20)	.61 (.21)
	[12.11, 17.83]	[12.94, 19.01]	[1.17, 1.78]	[.32, .98]
EN19	12.67 (1.50)	17.58 (1.89)	1.52 (.21)	.62 (.25)
	[10.55, 15.23]	[14.92, 20.85]	[1.22, 1.86]	[.31, 1.07]
EN20	11.92 (1.27)	17.48 (1.76)	1.53 (.23)	.48 (.24)
	[10.08, 14.01]	[14.95, 20.39]	[1.20, 1.90]	[.07, .81]
EN21	13.45 (1.74)	18.27 (2.10)	1.48 (.22)	.62 (.24)
	[11.21, 16.59]	[15.61, 22.06]	[1.15, 1.83]	[.29, 1.05]
EN22	15.39 (1.94)	15.74 (2.01)	1.52 (.22)	.58 (.21)
	[12.37, 18.45]	[12.61, 18.87]	[1.21, 1.88]	[.26, .92]
EN23	13.97 (1.69)	17.19 (1.89)	1.51 (.20)	.68 (.24)
	[11.60, 16.91]	[14.54, 20.52]	[1.20, 1.84]	[.37, 1.12]
EN24	10.99 (1.24)	17.81 (2.06)	1.47 (.25)	.51 (.24)
	[9.11, 13.01]	[14.80, 21.31]	[1.07, 1.83]	[.11, .87]
EN25	12.92 (1.55)	17.51 (1.84)	1.49 (.21)	.64 (.24)
	[10.77, 15.58]	[14.95, 20.70]	[1.16, 1.82]	[.33, 1.07]

**Table F.8:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 English listeners in the CL condition.

<b>Par.</b>	<b>intercept</b>	<b>VOT wt.</b>	<b>F0 wt.</b>	<b>tone wt.</b>
EC1	13.99 (2.72) [10.31, 18.75]	17.56 (3.11) [13.25, 22.99]	1.73 (.31) [1.34, 2.27]	.93 (.38) [.39, 1.60]
EC2	14.81 (2.69) [11.01, 19.50]	18.51 (3.07) [14.15, 23.71]	1.63 (.28) [1.24, 2.12]	.38 (.33) [-.14, .90]
EC3	11.90 (2.34) [8.49, 15.86]	17.50 (3.12) [12.85, 22.68]	1.72 (.34) [1.24, 2.31]	1.05 (.44) [.42, 1.80]
EC4	14.67 (2.56) [10.90, 19.03]	18.39 (2.93) [14.03, 23.22]	1.57 (.26) [1.17, 2.01]	.55 (.33) [.02, 1.10]
EC5	8.52 (1.50) [6.27, 11.04]	12.69 (2.17) [9.43, 16.23]	1.70 (.33) [1.24, 2.27]	1.04 (.40) [.45, 1.73]
EC6	9.25 (1.54) [7.00, 11.87]	13.43 (2.11) [10.31, 16.92]	1.52 (.31) [1.03, 2.03]	.09 (.38) [-.54, .69]
EC7	13.43 (2.96) [9.05, 18.43]	16.57 (3.47) [11.26, 22.27]	1.50 (.27) [1.06, 1.92]	.24 (.34) [-.32, .74]
EC8	14.26 (2.90) [10.02, 19.07]	16.62 (3.20) [11.88, 22.05]	1.57 (.25) [1.19, 1.98]	.74 (.33) [.25, 1.31]
EC9	4.81 (.73) [3.73, 6.06]	5.29 (.80) [4.10, 6.63]	1.28 (.26) [.85, 1.68]	.34 (.24) [-.04, .73]
EC10	15.75 (2.64) [11.93, 20.19]	21.13 (3.25) [16.37, 26.65]	1.73 (.33) [1.28, 2.30]	.83 (.43) [.20, 1.52]
EC11	14.90 (2.68) [11.03, 19.59]	17.37 (2.96) [13.18, 22.45]	1.65 (.27) [1.28, 2.12]	.65 (.31) [.17, 1.16]
EC12	15.28 (2.66) [11.54, 19.90]	19.51 (3.13) [15.00, 24.85]	1.68 (.29) [1.28, 2.19]	.73 (.37) [.18, 1.36]
EC13	13.15 (2.74) [9.32, 18.05]	15.23 (3.06) [10.96, 20.55]	1.54 (.25) [1.19, 1.96]	.34 (.29) [-.11, .80]
EC14	14.93 (2.78)	17.30 (3.05)	1.66 (.29)	.25 (.30)

**Table F.8 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 English listeners in the CL condition.

<b>Par.</b>	<b>intercept</b>	<b>VOT wt.</b>	<b>F0 wt.</b>	<b>tone wt.</b>
	[10.85, 19.65]	[12.86, 22.48]	[1.29, 2.15]	[−.24, .73]
EC15	14.37 (2.33)	19.92 (3.05)	1.77 (.33)	.93 (.42)
	[11.03, 18.25]	[15.54, 25.08]	[1.33, 2.36]	[.32, 1.65]
EC16	12.22 (1.85)	10.90 (1.68)	1.15 (.30)	−.10 (.30)
	[9.38, 15.33]	[8.35, 13.66]	[.64, 1.58]	[−.59, .38]
EC17	14.69 (2.73)	16.28 (2.97)	1.57 (.25)	.22 (.28)
	[10.83, 19.38]	[12.03, 21.32]	[1.21, 2.00]	[−.23, .66]
EC18	15.78 (2.55)	21.16 (3.14)	1.74 (.32)	.83 (.41)
	[12.00, 20.14]	[16.49, 26.56]	[1.31, 2.27]	[.21, 1.53]
EC19	11.65 (1.85)	9.18 (1.47)	1.15 (.29)	−.45 (.30)
	[8.83, 14.68]	[6.92, 11.57]	[.67, 1.59]	[−.93, .02]
EC20	15.39 (2.58)	15.21 (2.66)	1.36 (.26)	.23 (.30)
	[11.67, 19.91]	[11.42, 19.88]	[.94, 1.74]	[−.24, .72]
EC21	15.43 (2.77)	18.82 (3.15)	1.73 (.32)	.94 (.38)
	[11.42, 20.20]	[14.28, 24.17]	[1.32, 2.28]	[.40, 1.61]
EC22	13.07 (2.56)	16.17 (2.89)	1.61 (.27)	.31 (.32)
	[9.44, 17.59]	[12.03, 21.21]	[1.24, 2.06]	[−.22, .80]
EC23	14.92 (2.68)	17.98 (3.00)	1.68 (.28)	.42 (.31)
	[11.11, 19.59]	[13.65, 23.12]	[1.30, 2.18]	[−.07, .89]
EC24	15.20 (2.74)	19.68 (3.27)	1.64 (.31)	.68 (.39)
	[11.27, 19.82]	[14.93, 25.12]	[1.19, 2.15]	[.09, 1.32]
EC25	14.79 (2.32)	21.47 (3.30)	1.75 (.35)	.85 (.46)
	[11.41, 18.70]	[16.56, 27.02]	[1.24, 2.36]	[.15, 1.62]

**Table F.9:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 Mandarin listeners in the non-CL condition.

<b>Par.</b>	<b>intercept</b>	<b>VOT wt.</b>	<b>F0 wt.</b>	<b>tone wt.</b>
MN1	11.23 (1.95) [8.53, 14.62]	14.27 (2.28) [11.12, 18.21]	.88 (.33) [.43, 1.45]	.75 (.29) [.40, 1.28]
MN2	10.05 (1.59) [7.75, 12.74]	17.15 (3.12) [12.67, 22.46]	.24 (.33) [-.32, .72]	.48 (.23) [.10, .83]
MN3	11.58 (1.72) [9.08, 14.56]	17.87 (2.83) [13.80, 22.81]	.42 (.35) [-.14, .94]	.54 (.24) [.16, .92]
MN4	11.24 (1.77) [8.74, 14.28]	16.40 (2.50) [12.86, 20.73]	.57 (.33) [.07, 1.09]	.54 (.24) [.16, .93]
MN5	8.39 (1.49) [6.25, 10.93]	12.36 (1.98) [9.43, 15.63]	.74 (.27) [.35, 1.21]	.61 (.22) [.32, 1.00]
MN6	8.91 (1.44) [6.74, 11.33]	15.56 (2.87) [11.39, 20.57]	.55 (.29) [.10, 1.02]	.49 (.21) [.16, .83]
MN7	7.16 (1.44) [5.18, 9.58]	12.14 (2.81) [8.51, 16.97]	.71 (.27) [.31, 1.16]	.46 (.21) [.11, .76]
MN8	8.62 (1.59) [6.31, 11.40]	13.42 (2.54) [9.91, 17.78]	.42 (.31) [-.11, .85]	.51 (.21) [.16, .84]
MN9	6.89 (1.04) [5.32, 8.58]	10.58 (1.53) [8.30, 13.13]	.76 (.27) [.36, 1.19]	.64 (.22) [.36, 1.04]
MN10	6.18 (1.52) [4.13, 8.89]	12.02 (3.15) [8.01, 17.79]	.76 (.29) [.35, 1.27]	.47 (.20) [.15, .79]
MN11	11.71 (1.92) [9.05, 15.04]	16.29 (2.51) [12.72, 20.40]	.50 (.34) [-.06, 1.01]	.54 (.26) [.14, .96]
MN12	9.09 (1.53) [6.82, 11.70]	16.53 (3.13) [11.95, 22.03]	.21 (.31) [-.32, .68]	.33 (.25) [-.10, .67]
MN13	11.64 (1.79) [9.06, 14.71]	16.14 (2.30) [12.83, 20.03]	.56 (.32) [.06, 1.06]	.55 (.25) [.16, .93]
MN14	11.71 (1.89)	16.36 (2.57)	.45 (.34)	.53 (.25)

**Table F.9 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 Mandarin listeners in the non-CL condition.

<b>Par.</b>	<b>intercept</b>	<b>VOT wt.</b>	<b>F0 wt.</b>	<b>tone wt.</b>
	[8.98, 14.93]	[12.64, 20.82]	[−.11, .94]	[.13, .92]
MN15	11.79 (1.77)	18.17 (2.95)	.41 (.34)	.53 (.25)
	[9.21, 14.81]	[14.02, 23.29]	[−.17, .91]	[.13, .92]
MN16	9.72 (1.55)	16.91 (3.13)	.25 (.33)	.49 (.22)
	[7.49, 12.37]	[12.53, 22.25]	[−.31, .73]	[.16, .84]
MN17	12.59 (1.98)	18.88 (2.99)	.38 (.37)	.51 (.27)
	[9.74, 15.95]	[14.63, 24.07]	[−.24, .92]	[.09, .93]
MN18	9.59 (1.68)	12.15 (1.92)	.58 (.29)	.57 (.22)
	[7.16, 12.43]	[9.32, 15.35]	[.09, 1.01]	[.23, .91]
MN19	8.76 (1.44)	15.46 (2.88)	.63 (.28)	.53 (.21)
	[6.66, 11.21]	[11.46, 20.52]	[.21, 1.11]	[.22, .88]
MN20	6.83 (1.06)	7.81 (1.17)	1.03 (.28)	.74 (.23)
	[5.27, 8.60]	[6.11, 9.81]	[.61, 1.50]	[.43, 1.16]
MN21	7.36 (1.72)	9.60 (2.03)	.63 (.27)	.46 (.21)
	[5.05, 10.39]	[6.80, 13.11]	[.19, 1.04]	[.09, .76]
MN22	8.00 (1.60)	16.04 (3.43)	.23 (.29)	.33 (.23)
	[5.67, 10.74]	[11.11, 21.89]	[−.24, .68]	[−.07, .65]
MN23	10.07 (2.17)	12.03 (2.50)	.86 (.30)	.79 (.29)
	[7.06, 13.91]	[8.48, 16.45]	[.42, 1.37]	[.43, 1.32]
MN24	11.64 (1.82)	17.97 (2.93)	.46 (.34)	.54 (.25)
	[9.08, 14.77]	[13.89, 23.02]	[−.09, .96]	[.18, .95]
MN25	5.41 (.96)	6.10 (1.05)	1.00 (.28)	.61 (.21)
	[4.02, 7.14]	[4.59, 7.96]	[.60, 1.46]	[.30, .96]

**Table F.10:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 Mandarin listeners in the CL condition.

<b>Par.</b>	<b>intercept</b>	<b>VOT wt.</b>	<b>F0 wt.</b>	<b>tone wt.</b>
MC1	13.99 (2.72) [10.31, 18.75]	17.56 (3.11) [13.25, 22.99]	1.73 (.31) [1.34, 2.27]	.93 (.38) [.39, 1.60]
MC2	14.81 (2.69) [11.01, 19.50]	18.51 (3.07) [14.15, 23.71]	1.63 (.28) [1.24, 2.12]	.38 (.33) [-.14, .90]
MC3	11.90 (2.34) [8.49, 15.86]	17.50 (3.12) [12.85, 22.68]	1.72 (.34) [1.24, 2.31]	1.05 (.44) [.42, 1.80]
MC4	14.67 (2.56) [10.90, 19.03]	18.39 (2.93) [14.03, 23.22]	1.57 (.26) [1.17, 2.01]	.55 (.33) [.02, 1.10]
MC5	8.52 (1.50) [6.27, 11.04]	12.69 (2.17) [9.43, 16.23]	1.70 (.33) [1.24, 2.27]	1.04 (.40) [.45, 1.73]
MC6	9.25 (1.54) [7.00, 11.87]	13.43 (2.11) [10.31, 16.92]	1.52 (.31) [1.03, 2.03]	.09 (.38) [-.54, .69]
MC7	13.43 (2.96) [9.05, 18.43]	16.57 (3.47) [11.26, 22.27]	1.50 (.27) [1.06, 1.92]	.24 (.34) [-.32, .74]
MC8	14.26 (2.90) [10.02, 19.07]	16.62 (3.20) [11.88, 22.05]	1.57 (.25) [1.19, 1.98]	.74 (.33) [.25, 1.31]
MC9	4.81 (.73) [3.73, 6.06]	5.29 (.80) [4.10, 6.63]	1.28 (.26) [.85, 1.68]	.34 (.24) [-.04, .73]
MC10	15.75 (2.64) [11.93, 20.19]	21.13 (3.25) [16.37, 26.65]	1.73 (.33) [1.28, 2.30]	.83 (.43) [.20, 1.52]
MC11	14.90 (2.68) [11.03, 19.59]	17.37 (2.96) [13.18, 22.45]	1.65 (.27) [1.28, 2.12]	.65 (.31) [.17, 1.16]
MC12	15.28 (2.66) [11.54, 19.90]	19.51 (3.13) [15.00, 24.85]	1.68 (.29) [1.28, 2.19]	.73 (.37) [.18, 1.36]
MC13	13.15 (2.74) [9.32, 18.05]	15.23 (3.06) [10.96, 20.55]	1.54 (.25) [1.19, 1.96]	.34 (.29) [-.11, .80]
MC14	14.93 (2.78)	17.30 (3.05)	1.66 (.29)	.25 (.30)

**Table F.10 continued:** Summary of posterior distributions, in terms of mean (sd) [89% CrI], of perceptual weights along manipulated dimensions for individual L1 Mandarin listeners in the CL condition.

<b>Par.</b>	<b>intercept</b>	<b>VOT wt.</b>	<b>F0 wt.</b>	<b>tone wt.</b>
	[10.85, 19.65]	[12.86, 22.48]	[1.29, 2.15]	[−.24, .73]
MC15	14.37 (2.33)	19.92 (3.05)	1.77 (.33)	.93 (.42)
	[11.03, 18.25]	[15.54, 25.08]	[1.33, 2.36]	[.32, 1.65]
MC16	12.22 (1.85)	10.90 (1.68)	1.15 (.30)	−.10 (.30)
	[9.38, 15.33]	[8.35, 13.66]	[.64, 1.58]	[−.59, .38]
MC17	14.69 (2.73)	16.28 (2.97)	1.57 (.25)	.22 (.28)
	[10.83, 19.38]	[12.03, 21.32]	[1.21, 2.00]	[−.23, .66]
MC18	15.78 (2.55)	21.16 (3.14)	1.74 (.32)	.83 (.41)
	[12.00, 20.14]	[16.49, 26.56]	[1.31, 2.27]	[.21, 1.53]
MC19	11.65 (1.85)	9.18 (1.47)	1.15 (.29)	−.45 (.30)
	[8.83, 14.68]	[6.92, 11.57]	[.67, 1.59]	[−.93, .02]
MC20	15.39 (2.58)	15.21 (2.66)	1.36 (.26)	.23 (.30)
	[11.67, 19.91]	[11.42, 19.88]	[.94, 1.74]	[−.24, .72]
MC21	15.43 (2.77)	18.82 (3.15)	1.73 (.32)	.94 (.38)
	[11.42, 20.20]	[14.28, 24.17]	[1.32, 2.28]	[.40, 1.61]
MC22	13.07 (2.56)	16.17 (2.89)	1.61 (.27)	.31 (.32)
	[9.44, 17.59]	[12.03, 21.21]	[1.24, 2.06]	[−.22, .80]
MC23	14.92 (2.68)	17.98 (3.00)	1.68 (.28)	.42 (.31)
	[11.11, 19.59]	[13.65, 23.12]	[1.30, 2.18]	[−.07, .89]
MC24	15.20 (2.74)	19.68 (3.27)	1.64 (.31)	.68 (.39)
	[11.27, 19.82]	[14.93, 25.12]	[1.19, 2.15]	[.09, 1.32]
MC25	14.79 (2.32)	21.47 (3.30)	1.75 (.35)	.85 (.46)
	[11.41, 18.70]	[16.56, 27.02]	[1.24, 2.36]	[.15, 1.62]

## **Appendix G**

# **Supporting Materials for Chapter 6**

### **G.1 Statistical Model Specification**

Likelihood:

$$\begin{aligned}
& \stackrel{\ell=\text{Eng}}{g=\text{EL}}/p/ \text{ response}_i^j \sim \text{Bernoulli}( \\
& \quad \stackrel{\ell=\text{Eng}}{g=\text{EL}} P_{\text{guess}}^j \cdot 0.25 + \\
& \quad (1 - \stackrel{\ell=\text{Eng}}{g=\text{EL}} P_{\text{guess}}^j) \cdot \text{logit}^{-1}( \\
& \quad \quad (\beta_0 + g=\text{EL} w_0 + g=\text{EL} u_0^j) + \\
& \quad \quad (\beta_1 + g=\text{EL} w_1 + g=\text{EL} u_1^j) \cdot \text{VOT}_i + \quad \quad \quad (\text{VOT}) \\
& \quad \quad (\beta_2 + g=\text{EL} w_2 + g=\text{EL} u_2^j) \cdot \text{F0}_i + \quad \quad \quad (\text{F0}) \\
& \quad \quad (\beta_3 + g=\text{EL} w_3 + g=\text{EL} u_3^j) \cdot \text{TONE } 1_i \quad \quad \quad (\text{tone}) \\
& \quad )
\end{aligned}$$

$$\begin{aligned}
& \stackrel{\ell=\text{Eng}}{g=\text{ML}}/p/ \text{ response}_i^j \sim \text{Bernoulli}( \\
& \quad \stackrel{\ell=\text{Eng}}{g=\text{ML}} P_{\text{guess}}^j \times 0.25 + \\
& \quad (1 - \stackrel{\ell=\text{Eng}}{g=\text{ML}} P_{\text{guess}}^j) \times \text{logit}^{-1}( \\
& \quad \quad (\beta_0 + g=\text{ML} w_0 + \stackrel{\ell=\text{Eng}}{g=\text{ML}} v_0 + g=\text{ML} u_0^j) + \\
& \quad \quad (\beta_1 + g=\text{ML} w_1 + \stackrel{\ell=\text{Eng}}{g=\text{ML}} v_1 + g=\text{ML} u_1^j) \cdot \text{VOT}_i + \quad \quad \quad (\text{VOT}) \\
& \quad \quad (\beta_2 + g=\text{ML} w_2 + \stackrel{\ell=\text{Eng}}{g=\text{ML}} v_2 + g=\text{ML} u_2^j) \cdot \text{F0}_i + \quad \quad \quad (\text{F0}) \\
& \quad \quad (\beta_3 + g=\text{ML} w_3 + \stackrel{\ell=\text{Eng}}{g=\text{ML}} v_3 + g=\text{ML} u_3^j) \cdot \text{TONE } 1_i + \quad \quad \quad (\text{tone}) \\
& \quad )
\end{aligned}$$

$$\begin{aligned}
& \stackrel{\ell=\text{Man}}{g=\text{ML}}/p/ \text{ response}_i^j \sim \text{Bernoulli}( \\
& \quad \stackrel{\ell=\text{Man}}{g=\text{ML}} P_{\text{guess}}^j \times 0.25 + \\
& \quad (1 - \stackrel{\ell=\text{Man}}{g=\text{ML}} P_{\text{guess}}^j) \times \text{logit}^{-1}( \\
& \quad \quad (\beta_0 + g=\text{ML} w_0 + \stackrel{\ell=\text{Man}}{g=\text{ML}} v_4 + g=\text{ML} u_4^j) + \\
& \quad \quad (\beta_1 + g=\text{ML} w_1 + \stackrel{\ell=\text{Man}}{g=\text{ML}} v_5 + g=\text{ML} u_5^j) \cdot \text{VOT}_i + \quad \quad \quad (\text{VOT}) \\
& \quad \quad (\beta_2 + g=\text{ML} w_2 + \stackrel{\ell=\text{Man}}{g=\text{ML}} v_6 + g=\text{ML} u_6^j) \cdot \text{F0}_i + \quad \quad \quad (\text{F0}) \\
& \quad \quad (\beta_3 + g=\text{ML} w_3 + \stackrel{\ell=\text{Man}}{g=\text{ML}} v_7 + g=\text{ML} u_7^j) \cdot \text{TONE } 1_i + \quad \quad \quad (\text{tone}) \\
& \quad )
\end{aligned}$$

Priors:

$$\begin{matrix} \ell=\text{Eng} \\ g=\text{EL} \end{matrix} P_{\text{guess}}^j, \begin{matrix} \ell=\text{Eng} \\ g=\text{ML} \end{matrix} P_{\text{guess}}^j, \begin{matrix} \ell=\text{Man} \\ g=\text{ML} \end{matrix} P_{\text{guess}}^j \sim \text{Unif}(0, 1), j = 1, 2, \dots, 25$$

$$\beta_0 \sim \mathcal{N}(10, 5)$$

$$\beta_1 \sim \mathcal{N}(15, 5)$$

$$\beta_2, \beta_3 \sim \mathcal{N}(0, 5)$$

$$\begin{bmatrix} g^W_0 \\ g^W_1 \\ \vdots \\ g^W_3 \end{bmatrix} \sim \mathcal{N}(0, {}^1S), g = \text{EL}, \text{ML}$$

$$\begin{bmatrix} \begin{matrix} \ell=\text{Eng} \\ g=\text{ML} \end{matrix} v_0 \\ \begin{matrix} \ell=\text{Eng} \\ g=\text{ML} \end{matrix} v_1 \\ \vdots \\ \begin{matrix} \ell=\text{Man} \\ g=\text{ML} \end{matrix} v_7 \end{bmatrix} \sim \mathcal{N}(0, {}^2S)$$

$$\begin{bmatrix} g=\text{EL} u_0^j \\ g=\text{EL} u_1^j \\ \vdots \\ g=\text{EL} u_3^j \end{bmatrix} \sim \mathcal{N}(0, {}^3S), j = 1, 2, \dots, 25$$

$$\begin{bmatrix} g=\text{ML} u_0^j \\ g=\text{ML} u_1^j \\ \vdots \\ g=\text{ML} u_7^j \end{bmatrix} \sim \mathcal{N}(0, {}^4S), j = 1, 2, \dots, 25$$

$${}^1S = \begin{bmatrix} {}^1\sigma_0 & 0 & \dots & 0 \\ 0 & {}^1\sigma_1 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & {}^1\sigma_3 \end{bmatrix} \quad {}^1R = \begin{bmatrix} {}^1\sigma_0 & 0 & \dots & 0 \\ 0 & {}^1\sigma_1 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & {}^1\sigma_3 \end{bmatrix}$$

$${}^2S = \begin{bmatrix} {}^2\sigma_0 & 0 & \dots & 0 \\ 0 & {}^2\sigma_1 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & {}^2\sigma_7 \end{bmatrix} \quad {}^2R = \begin{bmatrix} {}^2\sigma_0 & 0 & \dots & 0 \\ 0 & {}^2\sigma_1 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & {}^2\sigma_7 \end{bmatrix}$$

$$\begin{aligned}
{}^3S &= \begin{bmatrix} {}^3\sigma_0 & 0 & \cdots & 0 \\ 0 & {}^3\sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & {}^3\sigma_3 \end{bmatrix} & {}^3R &= \begin{bmatrix} {}^3\sigma_0 & 0 & \cdots & 0 \\ 0 & {}^3\sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & {}^3\sigma_3 \end{bmatrix} \\
{}^4S &= \begin{bmatrix} {}^4\sigma_0 & 0 & \cdots & 0 \\ 0 & {}^4\sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & {}^4\sigma_7 \end{bmatrix} & {}^4R &= \begin{bmatrix} {}^4\sigma_0 & 0 & \cdots & 0 \\ 0 & {}^4\sigma_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & {}^4\sigma_7 \end{bmatrix}
\end{aligned}$$

$${}^1\sigma_0, {}^1\sigma_1, \dots, {}^1\sigma_3 \sim \text{Exponential}(1)$$

$${}^2\sigma_0, {}^2\sigma_1, \dots, {}^2\sigma_7 \sim \text{Exponential}(1)$$

$${}^3\sigma_0, {}^3\sigma_1, \dots, {}^3\sigma_3 \sim \text{Exponential}(1)$$

$${}^4\sigma_0, {}^4\sigma_1, \dots, {}^4\sigma_7 \sim \text{Exponential}(1)$$

$${}^1R \sim \text{LKJCorr}(1)$$

$${}^2R \sim \text{LKJCorr}(1)$$

$${}^3R \sim \text{LKJCorr}(1)$$

$${}^4R \sim \text{LKJCorr}(1)$$

## G.2 Statistical Model Output

**Table G.1:** Marginal posterior summaries for population-level parameters from the combined perception model.

Perceptual cue	Mean	SD	89% CrI	$p(\text{dir.})$
intercept <sub>EL</sub> (Eng.)	11.87	.75	[10.74, 13.16]	$p(\beta > 0) = 1.00$
intercept <sub>ML</sub> (Eng.)	10.08	.58	[9.17, 11.03]	$p(\beta > 0) = 1.00$
intercept <sub>ML</sub> (Man.)	9.94	.61	[8.99, 10.93]	$p(\beta > 0) = 1.00$
VOT <sub>EL</sub> (Eng.)	15.53	.86	[14.21, 16.95]	$p(\beta > 0) = 1.00$
VOT <sub>ML</sub> (Eng.)	15.37	.86	[14.03, 16.78]	$p(\beta > 0) = 1.00$
VOT <sub>ML</sub> (Man.)	15.04	.90	[13.61, 16.50]	$p(\beta > 0) = 1.00$
post-stop F0 <sub>EL</sub> (Eng.)	1.38	.13	[1.17, 1.59]	$p(\beta > 0) = 1.00$
post-stop F0 <sub>ML</sub> (Eng.)	.90	.16	[.65, 1.15]	$p(\beta > 0) = 1.00$
post-stop F0 <sub>ML</sub> (Man.)	.58	.15	[.34, .83]	$p(\beta > 0) = 1.00$
tone <sub>EL</sub> (Eng.)	.51	.10	[.34, .68]	$p(\beta > 0) = 1.00$
tone <sub>ML</sub> (Eng.)	.43	.12	[.24, .61]	$p(\beta > 0) = 1.00$
tone <sub>ML</sub> (Man.)	.55	.11	[.37, .73]	$p(\beta > 0) = 1.00$
intercept <sub>EL</sub> (Eng.) – intercept <sub>ML</sub> (Eng.)	1.79	.87	[.37, 3.13]	$p(\beta > 0) = .98$
intercept <sub>ML</sub> (Eng.) – intercept <sub>ML</sub> (Man.)	.14	.55	[–.70, 1.04]	$p(\beta > 0) = .59$
VOT <sub>EL</sub> (Eng.) – VOT <sub>ML</sub> (Eng.)	.16	.99	[–1.31, 1.90]	$p(\beta > 0) = .55$
VOT <sub>ML</sub> (Eng.) – VOT <sub>ML</sub> (Man.)	.33	.80	[–.87, 1.62]	$p(\beta > 0) = .66$
post-stop F0 <sub>EL</sub> (Eng.) – post-stop F0 <sub>ML</sub> (Eng.)	.48	.20	[.15, .81]	$p(\beta > 0) = .99$
post-stop F0 <sub>ML</sub> (Eng.) – post-stop F0 <sub>ML</sub> (Man.)	.31	.20	[–.02, .61]	$p(\beta > 0) = .95$
tone <sub>EL</sub> (Eng.) – tone <sub>ML</sub> (Eng.)	.08	.15	[–.14, .34]	$p(\beta > 0) = .70$
tone <sub>ML</sub> (Eng.) – tone <sub>ML</sub> (Man.)	–.12	.15	[–.37, .11]	$p(\beta < 0) = .78$