The structure of intraindividual variability in speech production and perception and the link between modalities

Daniel Patrick Schweizer

Thesis submitted to the University of Ottawa
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Linguistics

Department of Linguistics
Faculty of Arts
University of Ottawa

© Daniel Patrick Schweizer, Ottawa, Canada, 2026

Acknowledgements

As with any achievement, this dissertation was not built by my hands alone but is the conclusion of a long path shaped by the incredible support and guidance from many people.

J'ai sans doute eu une chance énorme d'avoir Marc Brunelle comme superviseur. Grâce à son expertise, j'ai pu explorer une grande diversité de sujets, apprendre une multitude de nouvelles techniques et partager ma passion pour les langues. Il m'a toujours encouragé et n'a jamais hésité à me pousser à raffiner mes idées et à dépasser mes propres attentes. Je lui suis très reconnaissant de m'avoir invité à collaborer à de nombreux projets de recherche, de m'avoir permis de choisir mon chemin et de m'avoir aidé à découvrir le français canadien, une langue qui occupe désormais une place importante dans ma vie.

I offer many thanks also to Suzy Ahn, who was always happy to discuss ideas, provide technical assistance and offer rigorous feedback throughout this entire process. It was a joy to collaborate on many different projects.

I am of course grateful to my evaluation committee and all my professors at the University of Ottawa whose classes and discussions continuously allowed me to discover new areas of linguistics.

Cette thèse n'aurait pas pu se matérialiser sans la bonne volonté de la communauté francophone de la région d'Ottawa-Gatineau. Je remercie tous ceux qui ont choisi de participer à cette étude.

I cannot fail to recognize that such an endeavor was only possible due to the unwavering presence of family. I am eternally grateful to my parents, Janis and Ken, for showing me that this path was feasible in the first place and for supporting me unconditionally throughout all the twists and turns it took to get here.

Ya que sin brújula uno se pierde, agradezco de manera infinita a José Andrés, confidente incansable y montaña inquebrantable — ¡se pudo!

Y por último, cómo no mencionar a mi viajera favorita, Michela, quien a su manera también celebra este logro conmigo.

Table of Contents

Acknowledgements ii
Table of Contents iii
Abstract vi
Résumé viii
List of Tables x
List of Figures xiii
1. Introduction 1
1.1 Individual variability 2
1.1.1 Variability in production 2
1.1.2 Variability in perception 6
1.1.3 Sources of individual variability 10
1.2 Link between production and perception 13
1.2.1 Proposed theories 13
1.2.2 Empirical evidence of link between acoustic cues in production and perception 21
1.2.3 Empirical evidence of link between variability in production and perception 23
1.3 Research questions 25
1.4 Laurentian French 26
1.4.1 Stop voicing 28
1.4.2 Vowels 29
1.4.3 Fricatives 30
1.5 Experimental design 31
1.6 Outline of dissertation 35
2. Production experiment 36
2.1 Methodology 36
2.1.1 Participants 36
2.1.2 Stimuli 38
2.1.3 Procedure 41
2.1.4 Analysis of acoustic cues 42
2.1.5 Analysis of variability 44
2.1.6 Statistical analysis 45
2.2 Results 47

2.2.1 Voiced and voiceless stops 47
2.2.1.1 Proportion of voicing 51
2.2.1.2 Post-release aspiration duration 53
2.2.1.3 Variability 56
2.2.2 Vowels 63
2.2.2.1 Vowel formants (F1 and F2) 64
2.2.2.2 Variability 67
2.2.3 Fricatives 72
2.2.3.1 Spectral moments 72
2.2.3.2 Variability 74
2.2.4 Variability across segments 76
2.3 Interim summary and discussion 78
2.4 Inter-session variability (Production A and Production B) 85
2.4.1 Voiced and voiceless stops 85
2.4.2 Vowels 88
2.4.3 Fricatives 91
2.4.4 Summary 93
2.5 Conclusion 94
3. Perception experiment 96
3.1 Methodology 96
3.1.1 Participants 96
3.1.2 Stimuli 97
3.1.2.1 Voiced and voiceless stops 98
3.1.2.2 Vowels 100
3.1.2.3 Fricatives 102
3.1.3 Procedure and analysis of response data 103
3.1.4 Analysis of variability 105
3.1.5 Statistical analysis 108
3.2 Results 109
3.2.1 Voiced and voiceless stops 109
3.2.2 Vowels 112
3.2.3 Fricatives 114
3.2.4 Variability across continua 117

3.3 Conclusion 125
4. Link between production and perception 126
4.1 Voiced and voiceless stops 126
4.1.1 Methodology 126
4.1.2 Results 127
4.2 Vowels 136
4.2.1 Methodology 136
4.2.2 Results 140
4.3 Fricatives 146
4.3.1 Methodology 146
4.3.2 Results 147
4.4 Link between variability 150
4.5 Influence of participant characteristics 152
4.6 Discussion and conclusion 156
5. Discussion 159
5.1 Individual variability in production 159
5.2 Individual variability in perception 166
5.3 Variability in the production-perception link 172
5.4 Implications for language models 175
5.5 Role of variability 179
6. Conclusion 185
References 189
Appendix: Coefficient tables 217

Abstract

Considerable intraindividual variability is consistently observed in both the acoustic properties of speech production and the phonological categorization of speech perception. Although much of this variability can be explained by linguistic or extralinguistic factors, less attention has been paid to the structure of such variability across identical trials. This dissertation provides evidence that this variability is a structured and consistent property of an individual's speech, largely organized across acoustic and articulatory targets and is found in both production and perception. Comparison of the results of this dissertation with those obtained in research on motor and cognitive functions suggests that this structure may not be specific to speech but rather represents more general patterns of an individual's behavior.

Forty-nine subjects participated in a single-session production and perception experiment investigating acoustic properties of voiced and voiceless stops, vowels, and fricatives in Laurentian French. Ten of these participants carried out the production experiment twice to determine the stability of their speech. The degree of acoustic variability in production differed considerably across individuals but was correlated within the speech of an individual across segments sharing similar acoustic values and articulatory configurations, such as voiced stops or high vowels. Variability in perception, measured by a two-alternative forced choice identification task, was also found to be correlated only across continua involving similar acoustic values and articulatory configurations. No correlation was observed between production and perception variability, implying that variability in these two speech dimensions has a certain degree of independence. The degree of similarity between cue weights and categorical boundaries in production and perception differed widely across individuals. For example, although many individuals who produced voiced stops with incomplete closure voicing also relied less on the voicing cue in perception, this was not always the case. Similarly, participants whose vowel distributions presented no overlap in production, did not necessarily have more categorical perception responses. Production and perception values were more similar in participants who responded with less consistency on the perception task. The strength of a production-perception

link is therefore likely dependent both on the task type and participants' attention to each task, with variability being a possible indicator of a certain task-related approach.

The findings of this dissertation underscore the importance of considering intraindividual variability in linguistic experiments as a window into the organizational structure of phonetic and phonological information in speech production and perception. Similar findings across distinct scientific disciplines support the view that speech is a specialized tool likely governed by the same biological structures underlying all aspects of human behavior. To enhance our understanding of speech, it is therefore encouraged to employ a broader, interdisciplinary approach and to focus on both community-level and individual patterns.

Résumé

La variabilité intraindividuelle est continuellement observée tant dans les propriétés acoustiques de la production de la parole que dans la catégorisation phonologique de la perception de la parole. Bien qu'une grande partie de cette variabilité puisse être expliquée par des facteurs linguistiques ou extra-linguistiques, moins d'attention a été portée à la structure de cette variabilité dans des tests identiques. Cette thèse amène de nouveaux éléments de preuve que cette variabilité est une propriété structurée et constante de la parole d'un individu, organisée principalement par des cibles acoustiques et articulatoires présentes dans la production et la perception. La comparaison des constats de cette thèse avec des résultats dans des recherche sur les fonctions motrices et cognitives suggère qu'il est possible que cette structure ne soit pas spécifique à la parole mais qu'elle reflète plutôt des tendances générales du comportement d'un individu.

Quarante-neuf sujets ont participé à des expériences de production et perception menées en une seule séance qui ont sondé les propriétés acoustiques des occlusives voisées et non-voisées, des voyelles, et des fricatives du français laurentien. Dix de ces participants ont réalisé l'expérience de production deux fois afin de déterminer la stabilité de leur parole. Le degré de variabilité acoustique dans l'expérience de production diffère de manière considérable entre les individus mais est corrélée dans la parole d'un même individu entre des segments qui partagent des valeurs acoustiques ou des configurations articulatoires similaires, tels que les occlusives voisées ou les voyelles hautes. La variabilité perceptive, mesurée par une tâche d'identification forcée à deux alternatives, est également corrélée uniquement entre des continua qui impliquent des valeurs acoustiques et des configurations articulatoires similaires. Aucune corrélation n'a été observée entre la variabilité dans la production et la perception, ce qui implique que la variabilité dans ces deux modalités présente un certain degré d'indépendance. Le degré de similarité entre les poids acoustiques et les frontières catégorielles de la production et la perception diffère grandement entre les individus. Par exemple, bien que beaucoup des participants qui ont produit des occlusives voisées avec un voisement incomplet aient moins utilisé le voisement durant la perception, ce n'a pas été le cas de tous. De même, les participants dont les distributions

vocaliques ne présentaient aucun chevauchement en production n'avaient pas nécessairement une perception plus catégorielle. Les performances en production et en perception sont plus similaires chez les participants qui ont répondu avec moins de constance à la tâche de perception. La magnitude du lien entre la production et la perception dépend donc vraisemblablement tant du type de tâche que de l'attention accordée à cette tâche par l'individu, et la variabilité est un indicateur possible de l'approche choisie.

Les conclusions de cette thèse soulignent l'importance de considérer la variabilité intraindividuelle dans les expériences linguistiques comme une fenêtre sur la structure organisationnelle de l'information phonétique et phonologique dans la production et la perception de la parole. Des résultats similaires observés dans différents domaines scientifiques soutiennent l'argument que le langage est un outil spécialisé régi par les mêmes structures biologiques qui sous-tendent les autres aspects du comportement humain. Afin d'améliorer notre compréhension de la parole, il est donc recommandé d'adopter une approche plus large et interdisciplinaire, considérant les tendances observables aussi bien au niveau de la communauté qu'à celui de l'individu.

List of Tables

Table 1.1 F1-F2 distance in Hertz of the four vowels of interest in the current dissertation, reported in three previous studies on Laurentian French. 30

Table 2.1 Counts of native languages, other than Laurentian French, and L2s spoken by the 49 participants whose data was analyzed. 38

Table 2.2 Stimuli produced by participants in the production experiment. All second-syllable onsets and word-final vowels are target segments. Grey-shaded boxes represent words for which acoustic measurements were also taken for the word-initial stop. 40

Table 2.3 Additional stimuli in the production experiment where acoustic measurements were taken for the word-initial stops and the word-final vowels. 41

Table 2.4 Number of tokens produced by each participant for each segment of interest. Three participants produced four repetitions instead of five and thus have slightly lower counts. 41

Table 2.5 Pearson correlations between the Coefficient of Variation of five acoustic measures. A single asterisk indicates a significance of p = .01. *AspDur* indicates post-release aspiration duration and *VoiProp* indicates voicing proportion. 78

Table 2.6 Pearson correlations of the Coefficients of Variation of f0 across four vowels following voiced and voiceless stops. All correlations are highly significant, indicated by the asterisk (p < .000001). *Vcd* indicates the vowel was preceded by a voiced stop and *vls* indicates the vowel was preceded by a voiceless stop 81

Table 2.7 Pearson correlations between Coefficients of Variation of F1 or F2 of the four main vowels in the production experiment (/e, ɛ, o, u/) and the additional vowels /y/ and /a/. Significant correlations are indicated with one (p = .02) or two (p < .001) asterisks 83

Table 3.1 Stimuli used in the perception experiment for the three phonological categories tested. 97

Table 3.2 Duration of steps of closure voicing and post-release aspiration used to build the stop voicing continuum in the perception experiment. 99

Table 3.3 Composition of two blocks in the two-alternative forced choice perception experiment. Within each block, each token was heard four times. Each block was repeated twice for a total of eight repetitions per token. Overall, 504 participant responses were collected. 104

Table 3.4 Correlation coefficients and p-values between Coefficient of Unalikeability (CoU) values of all combinations of stop voicing continuum categories. 111

Table 3.5 Levels and counts of sociodemographic and linguistic variables. Categorical variables are written in italics and values in parentheses represent subject counts. Numerical variables are marked in bold with ranges are given in parentheses 120

Table 4.1 Pearson correlation coefficients between production-perception differences for the five categories tested. No correlation was significant (p > .2). 150

Table 4.2 Pearson correlation coefficients and p-values between Coefficients of Unalikeability (CoU) values representing intraindividual variability in perception and Coefficients of Variation (CoV) values representing intraindividual variability in production. 151

Table 4.3 Levels and counts of sociodemographic and linguistic variables. Categorical variables are written in italics and values in parentheses represent subject counts. Numerical variables are marked in bold with ranges are given in parentheses.....153

Table A1. Results of linear regression on closure duration for voiced and voiceless stops. 217

Table A2. Results of linear regression on voicing proportion for voiced stops. 217

Table A3. Results of linear regression on voicing proportion for voiceless stops. 218

Table A4. Results of linear regression on voicing proportion for voiced stops in word-medial position. 218

Table A5. Results of linear regression on voicing proportion for voiceless stops in word-medial position. 218

Table A6. Results of linear regression on post-release aspiration duration voiceless stops.....219

Table A7. Results of linear regression on post-release aspiration duration for partially voiced *voiced* stops. 219

Table A8. Results of linear regression on post-release aspiration duration for partially voiced *voiced* stops and all voiceless stops. 220

Table A9. Results of linear regression on Coefficient of Variation of post-release aspiration duration in voiceless stops. 221

Table A10. Results of linear regression on Coefficient of Variation of voicing proportion in voiced stops. 221

Table A11. Results of linear regression on Coefficient of Variation for F1 in vowels.....221

Table A12. Results of linear regression on Coefficient of Variation for F2 in vowels.....221

Table A13. Results of linear regression on type of Coefficient of Variation (F1 and F2) in vowels. 222

Table A14. Results of linear regression on normalized F1 of vowels across the two recording sessions (Production A and Production B).....222

Table A15. Results of linear regression on normalized F2 of vowels across the two recording sessions (Production A and Production B).....222

Table A16. Results of linear regression on Euclidean Distance between vowels across the two recording sessions (Production A and Production B). 223

Table A17. Results of linear regression on Coefficient of Variation of F1 of vowels across the two recording sessions (Production A and Production B). 223

Table A18. Results of linear regression on Coefficient of Variation of F2 of vowels across the two recording sessions (Production A and Production B). 223

Table A19. Results of linear regression on normalized Center of Gravity of fricatives across the two recording sessions (Production A and Production B). 223

Table A20. Results of linear regression on Coefficient of Variation of Center of Gravity of fricatives across the two recording sessions (Production A and Production B). 224

Table A21. Results of logistical regression on likelihood of VOICED responses for the stop voicing continuum in the perception experiment. 224

Table A22. Results of linear regression on Coefficient of Unalikeability across the five continua in the perception experiment. 224

List of Figures

Figure 1.1 Outline of experimental sessions and data collection. A production experiment (A) was carried out with ten participants. Three to five months later, these same participants returned to participate in a single-session, back-to-back production (B) and perception experiment. Production B was identical to Production A. Forty-eight additional participants carried out the single-session Production B and perception experiment. Of these 48 additional participants, nine were excluded, as explained in Section 2.1.1. 32

Figure 2.1 Example annotation of target word *daté* /date/ 'dated' with the following annotation labels: *pv* is the previous vowel /y/, *cl* is the closure of /d/, *pv2* is the vowel /a/, *cl2* is the closure of /t/, *op* is the final vowel /e/, *ov* is the onset of voicing of /d/, *cv* is the cessation of voicing of /d/ during stop closure, *rv* is the resumption of voicing of /d/ during the post-release /a/ and *ov2* is the onset of voicing of the final vowel /e/. 43

Figure 2.2 Closure duration of voiced and voiceless stops by segment position and place of articulation. 47

Figure 2.3 Waveforms of the five observed stop voicing patterns: (A) complete closure voicing with no post-release aspiration, (B) complete closure voicing with post-release aspiration, (C) wane (>50%) closure voicing with post-release aspiration, (D) carryover (<50%) closure voicing with post-release aspiration, (E) voiceless closure with post-release aspiration. 48

Figure 2.4 Model waveforms of two stop voicing patterns that were not observed in the production data: (1) partial closure voicing that ceases before release without post-release aspiration, (2) partial closure voicing that begins at closure midpoint and continues until the release. 50

Figure 2.5 Counts of voiced and voiceless stops produced with different voicing proportions in closure. 52

Figure 2.6 Density plot of post-release aspiration duration (ms) of all productions of voiceless stops. The Vowel /a/ is word-initial and the remaining vowels are word-medial. 54

FIgure 2.7 Density plot of post-release aspiration duration (ms) of all productions of voiced stops realized with incomplete closure voicing. The vowels /a/ and /ɑ/ are word-initial and the remaining vowels are in word-medial position. 55

Figure 2.8 Proportion of voiced stops produced with full closure voicing either with or without post-release aspiration, corresponding to the voicing patterns (B) and (A), respectively, as defined in Figure 2.3. 56

Figure 2.9 Boxplots of distributions of voicing proportion for voiced and voiceless stops, arranged left to right by the overall prevalence of partial voicing. Participants on the far left produced nearly all voiced stops with full closure voicing and all voiceless stops with no closure voicing. Participants on the far right either produced a large proportion of voiced stops with incomplete closure voicing or a large proportion of voiceless stops with carryover voicing. Note that the single dashes at 0% and 100% (for subjects 41 and 32, for example) are compressed boxes. 57

Figure 2.10 Pearson correlations between mean voicing proportion of voiced or voiceless stops across three places of articulation. 58

Figure 2.11 Pearson correlations between mean values of post-release aspiration duration of voiceless stops across three places of articulation. 58

Figure 2.12 Distribution of Coefficients of Variation (CoV), representing intraindividual variability, of voicing proportion of voiced stops (left panel) and post-release aspiration duration of voiceless stops (right panel) across speakers. 60

Figure 2.13 Pearson correlations of Coefficient of Variation (CoV) values of voicing proportion of voiced stops (top panel) and post-release aspiration duration of voiceless stops (bottom panel) across different places of articulation. 62

Figure 2.14 Scatterplot of normalized F1 and F2 values, representing the overall vowel space for the four vowels recorded in Production B. 65

Figure 2.15 Boxplots of normalized F1 and F2 values in Production B. 65

Figure 2.16 Individual speaker Pillai scores for the two vowel contrasts studied. A lower Pillai score indicates more overlap in vowel categories. Only statistically significant Pillai scores are shown. 66

Figure 2.17 Distribution of Coefficients of Variation (CoV), representing intraindividual variability, of F1 and F2 values across all speakers. 68

Figure 2.18 Significant Pearson correlations between the Coefficient of Variation of F1 and the Coefficient of Variation of F2 within the same vowel. 69

Figure 2.19 Significant Pearson correlation between the Coefficients of Variation of F1 of the vowel pairs /e-o/ and /ɛ-o/. 70

Figure 2.20 Pearson correlations between the Coefficients of Variation of F2 of distinct vowel pairs. 71

Figure 2.21 Density plots of normalized values for the first four spectral moments of the fricatives /s/ and /ʃ/. 73

Figure 2.22 Speaker-level Cohen's d values for the first four spectral moments of /s/ and /ʃ/, ordered from lowest (leftmost) to highest (rightmost) Cohen's d for COG. 74

Figure 2.23 Distribution of Coefficients of Variation (CoV), representing intraindividual variability, of Center of Gravity (COG) of fricatives across all speakers. 75

Figure 2.24 Pearson correlation between the Coefficient of Variation of COG in /s/ and /ʃ/. 76

Figure 2.25 Individual mean Coefficient of Variation (CoV) values for each relevant acoustic measure, arranged by the total sum of CoV measures (lowest to highest from left to right). 77

Figure 2.26 Significant Pearson correlations between the Coefficients of Variation of F1 in the vowel pair /y-u/ (left) and between the Coefficients of Variation of F2 in the vowel pair /a-ɛ/ (right). 81

Figure 2.27 Boxplots of normalized F1 and F2 values of the four main vowels of interest in this study (/e, ɛ, o, u/) as well as two additional vowels (/a/ and /y/) whose variability is analyzed in this section. 82

Figure 2.28 Voicing proportion for voiced and voiceless stops across both recording sessions (Production A and Production B) across all ten speakers. 86

Figure 2.29 Post-release aspiration duration for voiceless stops in both recording sessions (Production A and Production B) across all ten speakers. 87

Figure 2.30 Coefficients of Variation (CoV) of voicing proportion (left) and post-release aspiration duration (right) for voiced and voiceless stops, respectively, across the ten speakers who participated in both production sessions (A and B). 88

Figure 2.31 Normalized mean F1 and F2 values across production sessions A and B. The vowel symbols connected by a line correspond to those from the same speaker. 89

Figure 2.32 Euclidean distances, indicating the change in mean vocalic formant values, between Production A and Production B. 90

Figure 2.33 Coefficients of Variation (CoV) of F1 and F2 for the four vowels of interest across the ten speakers who participated in both production sessions (A and B). 91

Figure 2.34 Boxplots of the first four spectral moments of the dental (/s/) and alveopalatal (/ʃ/) fricative across production sessions. 92

Figure 2.35 Coefficients of Variation (CoV) of COG for the dental (/s/) and alveopalatal (/ʃ/) fricative across the ten speakers who participated in both production sessions (A and B). 93

Figure 3.1 Oscillogram and spectrogram of an example stimulus *la belle* (only /ab-pɛ/ shown). The solid red lines encompass 75 ms of closure voicing, corresponding to Step 3V. The dashed red line indicates the start of the 10ms of voicing wane extracted from a token in which voicing disappeared during the closure. The dashed blue lines encompass 35 ms of aspiration, corresponding to Step 3A. 100

Figure 3.2 F1 and F2 values of the five steps of the formant continuum between /e/ and /ɛ/ following /f/. Step 1 (in blue) corresponds to /ɛ/ and Step 5 (in red) corresponds to /e/. 101

Figure 3.3 F1 and F2 values of the five steps of the formant continuum between /o/ and /u/ following /t/. Step 1 (in blue) corresponds to /o/ and Step 5 (in red) corresponds to /u/. 101

Figure 3.4 Spectra of the five steps of the fricative continuum. Step 1 (in blue) corresponds to /ʃ/ and Step 5 (in red) corresponds to /s/. 103

Figure 3.5 Example presentation of stimuli in the two-alternative forced choice perception experiment conducted with OpenSesame (Mathôt et al. 2012). 104

Figure 3.6 Visualization of the calculation of boundary width by means of a quasibinomial logistic function for a participant with a shallower slope (36) and a steep slope (38). 106

Figure 3.7 Significant Pearson correlations between individual values of boundary width and Coefficient of Unalikeability. 108

Figure 3.8 Proportion of VOICED responses for the stop voicing contrasts in the perception experiment. Thick, dark lines represent the mean across all participants, while pale lines show individual trajectories. 109

Figure 3.9 Distribution of values of Coefficient of Unalikeability (CoU) across speakers, indicating individual response consistency across different trials for the four stop voicing continua in the perception experiment. A low CoU indicates high consistency. 111

Figure 3.10 Significant Pearson correlation between mean the Coefficient of Unalikeability (CoU) of velar stops and the mean CoU of bilabial stops in the perception experiment. 112

Figure 3.11 Proportion of /ɛ/ or /o/ responses for the two vowel continua in the perception experiment. Thick, dark lines represent the mean across all participants, while pale lines show individual trajectories. 113

Figure 3.12 Distribution of values of Coefficient of Unalikeability (CoU), indicating individual response consistency across trials for the two vowel continua in the perception experiment. A low CoU indicates high consistency. 113

Figure 3.13 Proportion of /ʃ/ responses for the fricative continuum in the perception experiment. Thick, dark lines represent the mean across all participants, while pale lines show individual trajectories. Participants were divided into three categories based on their response patterns. 114

Figure 3.14 Distribution of values of Coefficient of Unalikeability (CoU), indicating individual response consistency across different trials for the fricative continuum in the perception experiment. A low CoU indicates high consistency. 115

Figure 3.15 Coefficients of Unalikeability (CoU) for each of the perception continua tested, arranged by participants with lowest (leftmost) to highest (rightmost) overall variability. 118

Figure 3.16 Pearson correlations between Coefficients of Unalikeability (CoU) of the four perception continua and the age of L2 acquisition. 121

Figure 3.17 Pearson correlations between mean reaction times (ms) of the four perception continua and the corresponding Coefficient of Unalikeability. 123

Figure 4.1 Individual by-subject random slope (beta) coefficients extracted from a mixed-effect logistic regression model, representing the likelihood of choosing the VOICED category as either voicing proportion or aspiration duration increased. Subjects are arranged from lowest (leftmost) to greatest (rightmost) beta coefficient of voicing proportion. 128

Figure 4.2 Individual Cohen's ds by subject. A positive value indicates that higher values of that measure are associated with phonologically voiced stops and a negative value indicates that higher values of that measure are associated with phonologically voiceless stops. Subjects are arranged from lowest (leftmost) to greatest (rightmost) Cohen's d of voicing proportion. 129

Figure 4.3 Individual log-transformed Cohen's ds by subject. A positive value indicates that higher values of that measure are associated with phonologically voiced stops and a negative value indicates that higher values of that measure are associated with phonologically voiceless stops. Subjects are arranged from lowest (leftmost) to greatest (rightmost) Cohen's d of voicing proportion. 130

Figure 4.4 Significant correlations between normalized (scaled to a range of 0-1) Cohen's ds, representative of production weights, and beta coefficients, representative of perception weights. 131

Figure 4.5 Two example participants, one with a strong production-perception link (25) and one with a weak production-perception link (35), for the acoustic property of voicing proportion.132

Figure 4.6 Two example participants, one with a strong production-perception link (10) and one with a weak production-perception link (2), for the acoustic property of post-release aspiration duration. 133

Figure 4.7 Production-perception link coefficients, arranged by degree of overall link across both acoustic properties, from those with the strongest (leftmost) to those with the weakest (rightmost). 134

Figure 4.8 Pearson correlations between mean Coefficients of Unalikeability (CoU), measuring response consistency on the stop voicing continuum in the perception task, and the production-perception link coefficient. A lower value on the x-axis indicates a stronger production-perception link. 135

Figure 4.9 Representation of the calculation of the perceptual midpoint for the /e-ɛ/ continuum for a sample participant. First, the boundary point, corresponding to 50% probability of a participant selecting either vowel category, was extracted, as indicated by the red circle in the lefthand plot. Then, the normalized F1 and F2 values of the perception stimulus located at this continuum step were taken as the perceptual midpoint, as indicated by the red circles in the righthand plot. The yellow circles represent the actual formant values of the model speaker extracted from the stimuli used in the perception experiment. Interpolated values were calculated to fill in the spaces between each step for participants whose boundary point did not fall directly on one of the five steps. 138

Figure 4.10 Representation of the calculation of the production midpoint for the two vowel pairs (/e-ɛ/ and /u-o/) for a sample participant. Red triangles show the centroid in normalized F2-F1 space of each vowel and purple diamonds represent the production midpoint between the two vowel distributions. 139

Figure 4.11 Production and perception midpoint values for all participants and both vowel contrasts. 141

Figure 4.12 Example participant with a weak production-perception link for both vowel continua. The left plot shows the response to both vowel continua in the perception task. The right plot shows formant values for all tokens in production, as well as the production (purple diamond) and perception (red circle) midpoint, corresponding to the theoretical boundary point between the two phonological categories. 142

Figure 4.13 Example participant with a strong production-perception link for both vowel continua. The left plot shows the response to both vowel continua in the perception task. The right plot shows formant values for all tokens in production, as well as the production (purple diamond) and perception (red circle) midpoint, corresponding to the theoretical boundary point between the two phonological categories. 143

Figure 4.14 Euclidean distances between production and perception boundary points, taken as the production-perception link coefficient. Values are arranged from the lowest (leftmost) to the highest (rightmost) combined sum across vowel pairs. 144

Figure 4.15 Pearson correlations between Coefficients of Unalikeability of the two vowel pairs tested in perception and the Euclidean distance between boundary points in production and perception, representative of the inter-modality link. 145

Figure 4.16 Production-perception link coefficients for the /s-ʃ/ fricative contrast, arranged from lowest (leftmost) to highest (rightmost). 148

Figure 4.17 Pearson correlation between the Coefficient of Unalikeability of fricative pair /s-ʃ/ in production and the fricative production-perception link coefficient. Solely for visualization purpose, the three participant groupings for this continuum, as explained in Section 3.2.3, are shown. 149

Figure 4.18 Pearson correlations between mean reaction times and production-perception differences for the four perception continua measured. The production-perception difference for the stop voicing continua represents only voicing proportion. 154

1. Introduction

Variability is an inherent and constant element of both speech production and speech perception. In production, the precise realization of a segment can present considerable acoustic differences across individual speakers and within the speech of a single speaker. *Interindividual variability*, where individuals differ from one another in their acoustic realization of phonological segments, can often be explained by sociolinguistic differences, such as dialect, age or sex (e.g., Babel & Munson 2014) or other directly identifiable properties of an individual, such as speech rate or emotional status (e.g., Gay 1968; Lee et al.2005). *Intraindividual variability*, where the same phonological segment is realized differently within the speech of an individual is often explained by linguistic factors, such as coarticulation, prosody, or semantics (e.g., Daniloff & Hammarberg 1973; Fougeron 1999). In both cases, however, there remains a non-negligible portion of variability that is *seemingly* random and unique to the individual (Whalen et al. 2018; Harper 2021). In perception, interindividual variability is seen in listeners' distinct categorizations of phonemes and weighting strategies of acoustic properties (e.g., Schertz 2014; Clayards 2018). In addition, intraindividual variability is also frequently observed, most often measured as response consistency across multiple repetitions of experimental tasks such as those probing phoneme identification and discrimination (e.g., McAllister Byun & Tiede 2017; Honda et al. 2024a).

Inter- and intraindividual variability is reported in nearly all phonetic studies, although it is often only mentioned in passing, and is accounted for by means of statistical models that allow for the emergence and analysis of general trends. Although such cross-subject patterns are undoubtedly important, patterns of individual differences in variability across acoustic and perceptual dimensions, as well as the relationship and possible correlations between these dimensions, can shed light on how language is structured and how this inherent variability may influence language use and change. Although less attention has been paid to inter- and intraindividual variability in the linguistics literature, the understanding of its importance is steadily increasing both in phonetics (Whalen et al. 2018; Chao et al. 2019; Cheng et al. 2021; Harper 2021) and in cognate fields, such as cognitive sciences (De Ribaupierre & Lecerf 2018;

Aschenbrenner & Jackson 2026) and motor sciences (Wu et al. 2014; Lyons et al. 2025). The importance of understanding inter- and intraindividual variability is also well-established in other fields, such as biology and ecology. For example, the characteristics of the microbiome, i.e. the combined composition of microorganisms within the human body, has been found to differ remarkably both across individuals (The Human Microbiome Project Consortium 2012) and within individuals at different points in time (Caporaso et al. 2011; Flores et al. 2014). Humans differ from one another not only in the taxonomic diversity of their microbiome but also in the magnitude of temporal variability. Most of this variability is not easily explained by subject characteristics, such as gender, age or body mass index, but appears either to be determined by environmental factors or to correspond to unique, personalized traits (The Human Microbiome Project Consortium 2012). Due to the high degree of interindividual variability in microbial makeup, personalized clinical applications are now being shown to enhance therapeutic efficacy across a range of technologies, such as probiotics (Montazeri-Najafabady 2025), chemotherapy and cardiovascular disease treatments (Schupack et al. 2022). The growing consensus is that inter- and intraindividual variability can provide not only information relating to the behavior or characteristics of a single individual but also offer insight into larger patterns relating to health, cognitive structures and human behavior in general.

1.1 Individual variability

1.1.1 Variability in production

While intraindividual variability is a consistent and inherent aspect of speech production, the question of whether it is specific to certain speech sounds or represents a more general, individual-level trait remains unresolved. Harper (2021) addressed this question by measuring intraindividual variability in acoustic and articulatory data of coronal English consonants taken from the Wisconsin XRMB database (Westbury 1994). Acoustic measurements included spectral moments for fricatives (/s/ and /ʃ/) and formants for liquids (/l/ and /ɹ/). A high degree of intraindividual variability in one segment did not necessarily indicate more variability in the speaker's speech overall, suggesting that the allowance for variability is segment-specific. Harper

(2021) interpreted these results as evidence that the cognitive representation of phonological units differs among speakers and that variability itself is part of a speaker's phonological grammar. Whalen et al. (2018) conducted a similar study utilizing the same XRMB database (Westbury 1994), looking at the production of nine English vowels. High rates of intraindividual variability were observed in the first three formants. In contrast to Harper (2021), they found that this variability *did* extend across segments. In other words, individual speakers produced all of their vowels with a comparable level of variability. It should be noted that these two studies looked at different types of phonological segments, namely consonants in Harper (2021) and vowels in Whalen et al. (2018). Given this and the differing results, it remains unclear whether intraindividual variability is specific to an individual, to a larger phonological class, such as vowels, or to a single phonological segment.

At the interindividual level, speakers also differ considerably from one another in the precise acoustic ranges and articulatory gestures used to produce speech sounds. This is the most consistently reported result in linguistic studies and it has been described for many different acoustic and articulatory measures. For example, considerable interindividual variability has been shown for the tongue position of coronal consonants in Kanada (Kochetov et al. 2018) and Arrernte (Tabain et al. 2020), in the production of voiced and voiceless stops in English (Ahn 2018a) and Laurentian French (Schweizer et al. 2025) and for coarticulatory nasalization in English (Beddor 2009). In acoustics, interindividual variability has been observed for phonetic voicing in English (Davidson 2016), Laurentian French (Schweizer et al. 2025) and Javanese (Brunelle et al. 2025), formant values in German, French and English (Strange et al. 2007), and fricative spectral cues in Russian (Ulrich et al. 2023) and English (Lo et al. 2023).

Such interindividual differences can also be observed in the relationship between articulation and acoustics within the speech of an individual. For example, Harper et al. (2020) observed through MRI data that American English speakers differ in the articulatory strategy they employ to produce the phoneme /ɹ/ and thus achieve the perceptually relevant low F3. Specifically, interindividual differences in the length and position of the palatal and pharyngeal constriction as well as the relationship between the two were observed. These interindividual differences were also observed at the level of articulatory-acoustic mappings, where the precise

effect of each constriction on lowering F3 depends on the relative importance each individual gives to these two constriction areas. Similarly, Blackwood Ximenes et al. (2017) observed that although articulatory data of English vowels, corresponding to tongue trajectories measured by means of Electromagnetic articulography (EMA), generally corresponded well to variability in F1 and F2 values, the magnitude of these correlations was highly speaker-specific. Such speaker-specific articulatory-acoustic mappings were also observed for the production of the German diphthong /ai/ (Weirich & Simpson 2018). Articulatory data was collected through the NDI Wave System which tracks the movement of sensors on precise points of the tongue, lips and jaw. Large interindividual differences in the degree of correlation between the F1-F2 difference and articulatory displacement during the production of the diphthong were observed. In other words, some speakers were found to produce a large change in F1-F2 distance with little corresponding articulatory movement and vice versa. In addition to highlighting individual differences, such results also suggest that the acoustic-articulatory relationship cannot be reduced to a single acoustic or articulatory indicator.

While interindividual variability most often refers to differences in acoustic ranges and articulatory gestures across speakers, it is also important to consider how individuals differ from one another in the degree of intraindividual variability. Harper (2021), for example, observed that the degree of intraindividual variability for both acoustic and articulatory dimensions varied considerably across participants. A similar observation was made by Ulrich et al. (2023) who analyzed 11 acoustic parameters in the production of eight Russian fricatives. The properties measured ranged from temporal, such as segment duration, and spectral, such as the center of gravity (COG) or kurtosis. Their results revealed significant interindividual differences in intraindividual variability overall, namely that some speakers' fricatives were produced with greater consistency, evidenced by lower standard deviations. They also found that the acoustic cue that displayed the greatest variability differed across individuals. Similar results have been found for the contrast of /s/ and /ʃ/ in English, namely that some speakers have consistently distinct categories while others produce both fricatives with considerable overlap (Newman et al. 2001; Lo et al. 2023).

Finally, production variability can be defined as the degree of stability of acoustic cues across time, recorded in distinct experimental sessions. Heald & Nusbaum (2015) recorded native English speakers producing isolated citation vowels in the morning, afternoon and evening on three different days, for a total of nine datasets. They found that there were small but systematic differences between the recordings at different times of day, but no significant change between days. Specifically, speakers' values of F1 increased by 13 Hz and values of f0 increased by approximately 9 Hz over the course of a day. They argued that the fact that F1 values did not significantly differ every morning shows that these values "reset" with sleep. Li et al. (2023) also observed relatively stable vowel productions across speech sessions in Norwegian speakers. Audibert & Fougeron (2022) investigated the speech of native European French speakers recorded across different days. Participants chose when to carry out each recording on their own, with a minimum of 24 hours between each session. They observed a large amount of inter- and intraindividual variability, with the most salient effects observed for mean f0, standard deviation of low frequency energy (between 0 and 1000 Hz), and mean harmonics to noise ratio (HNR). These acoustic properties all fall in lower frequencies and are arguably highly relevant to speech perception. Audibert & Fougeron (2022) also observed considerable differences in the degree of intraindividual variability, indicating that some speakers have quite stable acoustic cues, while others vary more.

As expected, all three types of variability, namely interindividual, intraindividual and inter-session, are commonly observed, although results are far from consistent and several open questions remain. While it is clear that individuals differ from one another in the degree of intraindividual variability within their speech, conflicting results have been reported regarding correlations in variability across segments. Additionally, the degree of stability of a speaker's acoustic realizations at different moments in time also appears to depend on both the individual and the acoustic cue considered. The observation of such inconsistent results underscores the importance of conducting a more systematic analysis of intraindividual variability to begin to understand what it may reveal about the structure of phonological information.

1.1.2 Variability in perception

As with variability in production, variability in perception is consistently observed in linguistic studies both at the interindividual and intraindividual level. One issue that complicates the comparison of perception variability across different studies is the use of a range of methodologies. While some researchers carry out categorization tasks, where listeners are instructed to assign a stimulus to one of the categories provided, others carry out discrimination tasks where listeners must identify the sound that differs from the other sounds presented in the same trial. While both methodologies undoubtedly provide important information, they arguably access different levels of speech perception and are both quite far removed from how speech is perceived in real-world contexts. Additionally, when probing perceptual categorization, two main task types are employed: a two-alternative forced choice (2AFC) and a visual analogue scaling (VAS) task. Results from these two task types do not always align with each other and likely also encourage participants to access distinct aspects of speech perception. It is important to review what has been observed across these different study types to gain a more complete picture of inter- and intraindividual variability in perception.

Regarding perceptual categorization, Hazan & Rosen (1991) presented participants with three types of stimuli to test place of articulation of English voiced stops in a 2AFC task: 1) covarying burst frequency and formant transitions, 2) burst frequency held constant and variable formant transitions and 3) formant transitions held constant and variable burst frequency. Little variability was observed for stimuli in the first group, where both acoustic cues were available, but participants varied considerably when presented with ambiguous stimuli in the 2) and 3) groups. Not only did participants differ widely from one another, but many individuals exhibited distinct behavioral responses to the 2) and 3) stimuli. Considerable interindividual variability on a 2AFC task has also been observed in the perception of relative and absolute duration for the contrast of Japanese singleton and geminate /t/ (Idemaru et al. 2012). In addition, a subset of Idemaru et al. (2012)'s participants repeated the same task three weeks after the first session and a high degree of consistency was observed, indicating stable perceptual cue weights.

Despite methodological differences, a similar degree of interindividual variability is also observed in categorization tasks that use the visual analogue scaling (VAS) task. In a VAS task, participants are instructed to position each token they hear on a sliding scale between the two elements of a minimal pair. Using this method, Kong & Edwards (2016) investigated the perception of VOT and f0 in the English voicing /t-da/ contrast. Participants carried out the same task twice in two separate experimental sessions timed a week apart. Despite large individual differences, participants could be grouped according to three perceptual strategies: categorical, intermediate and gradient. These patterns were found to be consistent across the two experimental sessions, providing evidence for more general speaker-specific perceptual strategies. Data from executive function tests, such as a Color-Word-Naming task, and non-verbal IQ tests, were not found to correlate with perceptual strategy, possibly indicating the uniqueness of speech perception among cognitive functions.

Interindividual variability is also commonly observed in discrimination tasks. For example, Perkell et al. (2004) observed a large degree of interindividual variability on an ABX discrimination task where participants chose whether the third stimuli heard (X) was identical to the first (A) or second (B) stimulus, for the English vowel contrast /ɑ-ʌ/. Franken et al. (2017) also observed considerable interindividual variability on a 4IAX perceptual discrimination task that probed listeners' perception of the Dutch vowel pairs /ɛ-ɪ/ and /ɑ-ɔ/. Participants carried out the same perception experiment in two separate blocks. In the 4IAX task, participants heard four acoustic stimuli, three of which were identical and one of which was unique, and had to then choose whether the second or third stimuli they heard was different from the rest. Although responses in both blocks were positively correlated, this correlation only explained around 40% of the variability in their responses, indicating that some participants were inconsistent in their responses even when hearing identical stimuli.

In addition to phonological categorization and discrimination, cue weights, i.e. the relative importance of acoustic cues, have been shown to vary considerably across individuals. Schertz et al. (2015) observed considerable interindividual variability in use of f0 and VOT in Korean speakers' perception of contrasts in their native Korean and non-native English. The weighting strategies for the non-native contrast were consistent across different experimental sessions.

Clayards (2018) tested listeners' weighting of primary and secondary acoustic cues across different phonological contrasts in English, such as stop voicing, vowels, and fricatives, and observed that despite large interindividual variability, the relative weighting of primary and secondary perceptual cues was consistent within a participant's responses.

Individuals not only differ from one another but can also shift their perceptual categories in order to adapt to distinct situations. Given the sheer variability that language users are exposed to throughout their lives, whether this stem from dialectal, generational or cultural differences, from the presence of an ongoing sound change, or simply from the uniqueness of different speakers, speech perception naturally needs to be more flexible than speech production. Indeed, listeners' awareness of certain sociophonetic characteristics has been shown to have the potential to greatly influence perception of identical segments. For example, speakers of American English from Detroit perceived more instances of Canadian Raising (tongue-fronting of the diphthongs /aɪ/ and /aʊ/) when they believed the speaker was Canadian (Niedzielski 1999) but less instances of Canadian Raising when told the same speaker was from Detroit. A similar result was found for speakers of New Zealand English who perceived the vowel /ɪ/ as being higher and fronter, a salient characteristic of Australian English, when the label 'Australian' was written on the answer sheet in front of them (Hay et al. 2006b). Other factors, such as recent exposure to a distinct dialect (Brunelle & Jannedy 2007; Hay et al. 2010) or the listener's attitude towards a dialect (Walker et al. 2018) have also been found to affect perception patterns of identical stimuli.

Flexibility in perceptual cues has been observed particularly in ongoing sound changes. In New Zealand English, there is an ongoing merger between the diphthongs /eə/ and /ɪə/, which is more advanced for younger speakers (Hay et al. 2006a). Listeners who make a distinction in production had lower discrimination accuracy when they assumed the speaker was younger as opposed to older, despite the stimuli being identical in both conditions (Hay et al. 2006a). Similar results of age-based sound changes altering the perception of identical tokens has been observed in other sound changes in New Zealand English (Drager 2011) and Houston English (Koops et al. 2008). Such speaker characteristics can also affect a listener's cue weights. For example, Seoul Korean listeners pay more attention to f0 when hearing a female voice compared to a male voice,

due to the more advanced f0-merger present in the speech of female speakers (Kong et al. 2011). Yu (2022) observed that English listeners have less reliance on VOT when the stimuli they hear are associated with a male face, as opposed to a female face, likely due to their knowledge that word-initial voiced and voiceless English stops are more similar in male speech (Robb et al. 2005).

As in production, variability is not only observed across individuals but also within the perception of an individual. In other words, participants may produce different responses even upon hearing identical stimuli. Such intraindividual variability, operationalized as response consistency, has been investigated primarily in categorization tasks, both using the 2AFC and VAS methodologies. For example, Kapnoula et al. (2017) and Honda et al. (2024a) measured English speakers' perception of distinct phonemic contrasts. Results from both task types in both studies revealed four distinct behavioral patterns in listeners. Gradient listeners were more attuned to subtle differences between similar stimuli, rating them differently on the VAS scale, whereas more categorical listeners preferred using the two extremes (0 or 100) in the rating task, producing very similar results as in the 2AFC task. These results again highlight the large differences between individuals in perception. These two types of listeners, however, could further be divided by their consistency, i.e. whether they chose the same response across trials for the same token. This measure of response consistency corresponds to the concept of intraindividual variability that will be explored in this dissertation. Honda et al. (2024a) also found that the degree of response consistency was similar across both VOT and vowel dimensions, suggesting that such behaviors may be unique traits of the individual and do not depend on a specific phonetic category.

Similar results were observed by Myers et al. (2024), who reported a moderate correlation between consistency measured by a VAS scale task in the identification of a /b-d/ continuum differing in F1 and F2 transitions, and of an /s-ʃ/ continuum, differing in peak frequency. Kim et al. (2025) also found that categorization consistency on a VAS task in English speakers was correlated across various continua for stop voicing, vowel identity and fricative identity. While such results seem to indicate that intraindividual variability in perception is a generalizable characteristic of the listener, this relationship is likely not entirely straightforward as some studies fail to find a correlation in consistency across continua. Fuhrmeister & Myers (2021), for

example, observed no correlation between response consistency on fricative and stop continua which varied across place of articulation. Similarly, Zaharchuk et al. (2026) observed significantly greater rates of inconsistency on a fricative /s-ʃ/ continuum than on a stop voicing or stop place continuum in both VAS and 2AFC tasks. Results such as these point to a more nuanced picture where response consistency may depend on the phonological contrast or acoustic information perceivers need to process. The observation of such contradictory results emphasizes that the nature of intraindividual variability in perception remains an unresolved matter.

Overall, the general trend observed across speech perception studies is that perception exhibits a high degree of individuality and flexibility. Participants differ not only in their categorization patterns but also in the consistency of these patterns across different phonological segments. It remains unclear if the degree of response consistency is a global trait affecting all speech segments or if it is dependent on the specific acoustic dimension or phonological category tested or the task itself. Additionally, the possible relationship between intraindividual production variability and response consistency on a perception task remains an important question to explore.

1.1.3 Sources of individual variability

While it is well-established that intraindividual variability is consistently observed in both speech production and speech perception, across a wide range of phonological and phonetic categories, the possible reasons behind the prevalence of such variability remain unclear. Regarding the articulation of speech sounds, interindividual differences can easily be attributed to physiological or anatomical variations. A common example is observed in the role the length of the vocal folds plays in determining an individual's average fundamental frequency (f0). While a speaker's f0 cannot be determined solely by vocal fold length, males tend to have a lower f0 range than females, which is in large part due to longer vocal folds (Zhang 2021). The shape of the palate, i.e., if it is flat or domed, has been shown to influence, but not entirely determine, an individual's articulatory strategy in producing coronal fricatives, such as /s/ or /ʃ/ in German (Weirich & Fuchs 2013). While such anatomical constraints have been shown to influence an

individual's production of speech sounds, they are crucially not entirely inevitable but appear rather to simply encourage a certain articulatory preference (Stone et al. 2012).

Given the observed influence that such morphological factors can have on interindividual variability, various studies have considered the influence they may have on intraindividual variability as well, although results are inconclusive. For example, while some studies have observed that individuals with flatter palates produce certain vowels with less intraindividual variability (Brunner et al. 2009; Rudy & Yunusova 2013), more recent studies have failed to find a significant correlation (Bakst 2021; Harper 2021). Such inconsistent results suggest that while anatomical differences may influence speech in some participants, it is likely only one of many factors that underlie the degree of intraindividual variability.

Reduced intraindividual variability in speech production has also been observed in individuals with heightened auditory sensitivity (Franken et al. 2017). Specifically, Franken et al. (2017) observed that individuals who responded with greater accuracy to a discrimination perception task also produced vowels with tighter distributions, which equates to less variable formant targets. Greater somatosensory acuity, measured by the ability to distinguish grooves on plastic domes with the tongue tip, has also been shown to correlate with greater distance between acoustic targets (Ghosh et al. 2010). Although acoustic distance does not directly reflect intraindividual variability, such results clearly show that individual differences in information processing can influence the realization of speech sounds.

Regarding inter- and intraindividual variability in perception, cognitive functions such as working memory and attentional switching ability have been shown to influence accurate phoneme discrimination (Ou et al. 2015; Ou & Law 2017). In addition, recent studies have determined that individual differences in the subcortical encoding of acoustic information can predict an individual's accuracy rate in a 2AFC categorization task, where listeners with greater categorization uncertainty, corresponding to response consistency, had less accurate neural encoding of the acoustic signal (Ou et al. 2022). This result was also found in a follow-up experiment testing the relative weight of vowel duration and F1 in an English vowel perception experiment. Specifically, the type of acoustic information encoded with greater accuracy was correlated with the more heavily weighted cue in the 2AFC perception task (Ou et al. 2023).

Honda et al. (2024b) also observed a similar, although non-significant, tendency whereby more consistent neural encoding of speech sounds was found in participants with more consistent responses on a VAS perception task. Such results suggest that interindividual differences in perception are at least partially due to unique neurocognitive differences in how the nervous system processes sounds. The observation that encoding accuracy was cue-independent also implies that different aspects of the acoustic signal are processed in distinct ways for each individual.

It is important to recognize the influence situational factors may have on the magnitude of intraindividual variability. A participant's emotional state, fatigue, level of distraction and even transient physiological conditions such as inflammation or discomfort can likely impact one's speech production and perception. Nevertheless, the fact that no variable explained above can consistently and accurately predict intraindividual variability, suggests that there may be some aspect inherent to each individual, stemming possibly from a combination of biological factors and learned behaviors, that determines one's rates of variability. In fact, such an idea of a unique, individual "fingerprint" of variability has also been proposed in motor sciences as an explanation for differing rates of intraindividual variability observed across distinct movement tasks (Jackson et al. 2020; Kasuga et al. 2022).

Finally, intraindividual variability in both production and perception can likely be partially attributed to the well-known observation that the brain's responses to identical sensory stimuli vary greatly (Stephani et al. 2020). This is because the precise response depends both on the external stimulus itself and on the brain's previous and current state. While this lack of consistency can often be seen as detrimental, Stephani et al. (2020) propose it reflects a necessary equilibrium between robustness and flexibility, which improves the brain's ability to process information. In sum, variability appears to be at the core of not only speech production and perception, but also larger neural processes that likely affect both speech modalities. At this point, no single underlying mechanism has been found that can predict an individual's rate of variability for a given sound or in a given context, although various external and internal factors can have a significant influence.

1.2 Link between production and perception

1.2.1 Proposed theories

A link between production and perception has been speculated upon and investigated for over seventy years. One of the main observations that has spurred on investigation in speech production and perception is the so-called "lack of invariance" (Casserly & Pilon 2010). This describes how it is difficult to connect the core idea of phonology that a language is composed of a discrete and limited set of phonemes with the observed range of variability present in normal speech. The exact acoustic realization of a phoneme can differ considerably between speakers and within the speech of an individual speaker, and the same phoneme can be realized with significant variability across different contexts. As outlined previously, much of this variability can be ascribed to sociolinguistic differences or phonetic and phonological phenomena, such as coarticulation or allophony, respectively. However, as is the focus of the current dissertation, a large amount of variability remains seemingly arbitrary and highly speaker-specific. Despite the abundant variability present in perceived speech, native speakers are generally able to arrive at the same overall phonological structure, arguably due to the formation of distinct categories of the same, or similar, abstract phonemes. A plethora of theoretical accounts have proposed a connection between production and perception that aim to explain the purported problem of "lack of invariance", although robust, neurophysiological evidence supporting such theories is hard to come by. In this section, a brief overview of the most influential models, their predictions and the empirical evidence supporting them will be reviewed. The goal of the current dissertation is not to argue for or against a specific speech model, but rather to encourage research in speech perception and production to consider the lack of invariance not as a problem but rather as an important window into the innerworkings of speech and open the door to future work on the role variability plays within the interplay of these two systems.

Developed in the 1960s, the main tenets of Motor Theory state that speech is processed through the perception of intended vocal tract gestures and that the motor system of speech is activated concurrently with perception (Liberman & Mattingly 1985). Listeners utilize the acoustic signal to internally reproduce the corresponding articulation and thus reach a conclusion

as to what the intended phoneme of the speaker was. According to this theory, variability in the acoustic signal is irrelevant if the differing patterns are all produced by the same articulatory configuration. Listeners need only understand the vocal tract configuration necessary to produce a specific sound to accurately perceive a phoneme and therefore essentially ignore subtle acoustic differences. For example, the second formant (F2) following the release of an English /d/ rises when the following vowel is /i/ but falls when the following vowel is /u/. These are two very distinct acoustic consequences that arise from arguably the same articulatory configuration of the voiced alveolar plosive, namely contact between the tongue tip and the alveolar ridge (Liberman et al. 1967). The Motor Theory claims that the fact that even when faced with such contrasting acoustic information, listeners are able to accurately and consistently perceive a /d/ is proof that what is perceived are gestures. It is important to note that this and most commonly cited examples of the so-called "lack of invariance" are in fact due to coarticulatory processes and are not representative of the seemingly random inter- and intraindividual variability of interest in this dissertation.

The fact that listeners' perception of speech in a noisy environment is increased if the speaker is visible, arguably allowing them to attend more closely to visual articulatory cues, is taken as support for the Motor Theory (Sumby & Pollack 1954; Whalen 1999). The McGurk effect is also often cited as indirect support for this theory due to the influence visual cues have on perception of identical acoustic signals (McGurk & MacDonald 1976). The McGurk effect designates a phenomenon in which listeners who hear one syllable, such as /pa/ but see a speaker producing a different syllable, such as /ka/, often hear a syllable located articulatorily in the middle between the velar and bilabial stop, i.e. /ta/ (McGurk & MacDonald 1976). If production gestures are simultaneously activated upon perception, it follows logically that this theory predicts a tight link between the two modalities, where speech perception is possible mainly by accessing knowledge of speech production.

Motor Theory naturally requires listeners to have a strong grasp of production in order to perceive speech. This tenet is challenged by observations of other phenomena in language acquisition and processing. Infants as young as two months old who not only are unable to produce linguistically meaningful speech but also lack the anatomical structures to accurately

articulate many speech sounds, are still able to distinguish between linguistic and non-linguistic sounds (Vouloumanos & Werker 2004). After only a few months from birth, infants are also able to correctly distinguish between different places of articulation and/or voiced and voiceless consonants, again without having the corresponding physical skill to produce these distinctions (Eimas et al. 1971; Werker & Tees 1984). Additional evidence against the Motor Theory is found in research on individuals who suffer from Broca's aphasia, i.e. damage to Broca's area in the motor cortex employed in speech production (Mohr et al. 1978; Bunker & Hills 2022). Despite an often reduced capability to produce speech, speech perception is largely unaffected (Moineau et al. 2005).

In the 1980s, an alternative theory, Direct Realism, was proposed, making a similar claim that speech is perceived and processed in physical gestures, but with the key difference that these gestures are directly recoverable from the acoustic signal. In other words, accessing the motor system is not necessary as the information is readily available as a physical, audible object (Fowler 1986; Fowler 1996; Fowler et al. 2003). Direct Realism claims that the gestures themselves are invariant, despite consequences of coarticulatory processes. Fowler (1996) claimed that perception across all dimensions, be it linguistic, visual or haptic, relies on the ability of the perceiver to infer the structure of the object that altered the medium through which the information was transferred from the outside world into the perceiver's mind. In other words, the actual medium itself is, in a way, ignored and is not considered a perceptual object, but the perceiver decodes the information contained within and directly *perceives* the shape or gestures that altered the medium in the first place. Such a theory relegates the acoustic signal and even in some way the auditory system to a mere transitional stage of perception, although substantial neurophysiological evidence to justify this assumption is lacking, just like for Motor Theory. Still, Direct Realism evades most of the critiques of Motor Theory in that it assumes that perception does not require brain regions responsible for production to be activated concurrently during speech perception. Evidence provided for this theory often includes the McGurk effect for the same reasons as described previously in support of the Motor Theory (McGurk & MacDonald 1976), although again it is suggested that such a phenomenon occurs because the *only* thing that listeners perceive in a direct manner are gestures. In a study by Porter & Lubker (1980), listeners

were instructed to shadow a model speaker and repeat the vowel they heard immediately after hearing it. Results showed that there was no significant difference in reaction time between the shadowing task and a simple production task, which arguably does not involve decision-making. Fowler (1996) takes this as strong evidence for Direct Realism, as she assumed that if speech perception relied exclusively on the auditory signal, such a fast reaction time would not be possible.

The approach that most obviously contrasts from Direct Realism and Motor Theory is the General Auditory approach that claims that speech perception relies not on articulatory gestures but rather exclusively on auditory cues that are accessed in the acoustic signal and subsequently mapped onto mental categories corresponding to phonemes (Diehl et al. 2004). Crucially, auditory approaches do not refer to gestures and therefore largely assume a separate system for production and perception. These systems are arguably connected through the abstract concept of phoneme, but neither is dependent on the other and they are not concurrently activated. One example of evidence brought forward to dispute the claim that gestures are the objects of speech perception is that different articulatory gestures that can lead to the same acoustic consequence, such as a rounding of the lips or lowering of the larynx, can result in a highly similar vocalic formant structure (Riordan 1977). A purely auditory approach, in which production and perception are completely separate mechanisms, is no longer supported by the literature, as many studies using cerebral imaging have since confirmed that certain regions of the motor cortex are activated upon perceiving speech (Pulvermüller et al. 2006; Meister et al. 2007; Tremblay & Small 2011; Cheung et al. 2016).

Support for a strict view of the Motor Theory was revived in the early 2000s with the discovery of mirror neurons (Rizzolatti & Arbib 1998; Galantucci et al. 2006). Mirror neurons are activated both when a physical movement is carried out as well as when an individual observes a conspecific carrying out the same action (Molenberghs et al. 2012; Bonini et al. 2022). This has been shown for various actions, for example when biting an apple or grabbing a ball and observing a video of someone else performing these actions (Buccino et al. 2001) or playing and listening to piano (Bangert et al. 2006). In the realm of linguistics, Pulvermüller et al. (2006) used fMRI to observe that the same brain regions were activated when participants produced nonce

syllables and when hearing these same syllables. Meister et al. (2007) observed that participants were unable to complete an otherwise simple phonetic discrimination task when transcranial magnetic stimulation was applied to the premotor cortex, leading them to claim that activation of the motor region is obligatory for accurate speech perception.

Despite this initial support for the Motor Theory, more recent studies have failed to find such a clearcut and consistent coactivation. Tremblay and Small (2011) observed in an fMRI study that the ventral premotor cortex was active in both production and perception, although stronger activation was observed in production. Additionally, differences in syllable complexity (a singleton or a consonant cluster in the onset) led to different activation levels in production but not in perception. Using high-density direct human cortical recordings, where electrodes are placed directly on the brain's surface, Cheung et al. (2016) observed that although the motor cortex was activated in both production and perception, the organization of these regions did not match up. That is, the regions associated with a specific place of articulation in production were not the same as those activated upon perception of the same place of articulation. In sum, most researchers currently do not support the idea that mirror neurons provide definitive proof of the strict version of the Motor Theory (Lotto et al. 2009; Holt & Lotto 2014; McGettigan & Tremblay 2018; Arjmandi & Behroozmand 2024). Although the research clearly indicates there is coactivation of production and perception on some level, this does not seem to be a requirement.

In the past 30 years, additional theories and approaches have been proposed that remove themselves slightly from the strong debate between gestural and auditory views. One example are exemplar-based theories that posit that production and perception targets are acquired through recognition of patterns of lexemes and individual phonemes, based on the probabilistic distribution of certain cues in the speech signal, leading to the creation of "exemplars" (Nearey 1997; Pierrehumbert et al. 2002; Pierrehumbert 2003). All exemplars are stored in one's memory and contain detailed information of every perceived and produced utterance of a specific word or phoneme. This implies that such theories inherently assume a shared system between production and perception. Variability, or the previously debated "lack of invariance", in an exemplar-based account is a core aspect of the speech system and language users actively make use of these probability distributions to accurately perceive and produce sounds. Such theories

can explain why listeners easily manage variability caused by coarticulation, a predictable process speakers have been exposed to since acquiring a language, but have more difficulty adapting to variability introduced by an unfamiliar dialect (Redford & Baese-Berk 2023). Exemplar theories also can easily justify lexical frequency effects, i.e. quicker reaction times in recognizing or producing more frequent words, by positing that there are simply more exemplars of such words stored in the mind (Hay et al. 2006a). This view therefore proposes an unequivocal link between production and perception although this link need not reflect a single individual's cue use but rather the overall tendency of the population in which the individual acquired the language (Schertz 2014).

One commonly mentioned downside of exemplar theories is the cognitive processing power thought to be needed to store and subsequently access every utterance during both speech production and perception (Baayen et al. 2013). Harper (2021) addressed this often-cited downside of exemplar theory by proposing a modified version of Dynamic Field Theory (Schöner et al. 1997) that represents phonological space as being composed of discrete units that correspond not to a single articulation but rather a wider target space for each segment, thus providing an allowance of variability in target selection. In other words, Harper (2021) proposed that individual utterances are stored as exemplars *and* that abstract representations are then formed based on these exemplar distributions, thus incorporating the degree of individual variability directly into phonological grammar.

An aspect of phonetic research that offers convincing evidence of a direct link between production and perception and is generally explained by means of exemplar theories is phonetic imitation or convergence, where one speaker alters their acoustic output to match that of another individual (Goldinger 1998; Shockley et al. 2004; Pardo 2006; Nielsen 2011, among others). Goldinger (1998) and Shockley et al. (2004) instructed participants to imitate each word they heard and subsequently compared the imitated utterance to a baseline utterance in an AXB perceptual task with other listener participants who accurately identified the imitation effect. Pardo (2006) recorded interactions between two speakers, without providing specific instructions to imitate speech, and discovered evidence of phonetic convergence by means of

perceptual assessments. In these studies, imitation was measured by external listener judgement and thus details regarding which acoustic properties were imitated are not provided.

Other imitation studies have focused more on specific acoustic characteristics to see how production and perception may be connected at a sub-phonemic level. Nielsen (2011) observed that speakers consistently replicated exceedingly long VOT values in voiceless, aspirated English stops without any explicit instruction to imitate. This VOT-lengthening was observed for initial stops of novel words that were not part of the listening task and extended across different places of articulation, suggesting that imitation affects an abstract level of representation, rather than a specific word or phoneme. A similar effect has been observed in vowel imitation, measured by F1-F2 distance (Babel 2012). Zellou et al. (2016) observed that participants imitated the degree of coarticulatory nasality in English words. This imitation was greater when exposed to increased nasality rather than decreased nasality, which was argued to be due to the facilitative effect coarticulatory nasalization has on successful word recognition. Zellou et al. (2016)'s result is similar to Nielsen (2011)'s study, in which no significant imitation was found for tokens with reduced VOT, and suggests that participants are less likely to imitate a property that may hamper communication.

Kim & Clayards (2019) observed that participants who were more perceptually sensitive to differences in vowel duration more accurately imitated this same property, offering potential evidence for an individual link between specific phonetic cues in production and perception. However, they found no evidence of such a link between baseline perception and production coefficients prior to the imitation task, indicating that the presence of such a link may be due to the imitation task itself. Schertz et al. (2023) observed that participants who scored higher on an accent discrimination task were also perceived as better accent imitators when explicitly told to imitate accents that differed in VOT. They argue that imitation is dependent on both a semi-automatic, generalized tendency towards phonetic convergence and individual, speaker-level factors, again hinting towards a certain degree of individual variability in such a production-perception link, at least with regards to imitation.

Several other models have been proposed to demonstrate and explain the link between production and perception, two of which will be briefly mentioned here. The model WEAVER++

(Levelt 1993, Levelt et al. 1999) assumes that speakers access perceptual information during production by means of an internal emulator, making use of both articulatory gestures and auditory cues. In addition to listening to one's own auditory signal produced while speaking, speakers simultaneously articulate the planned string of phonemes by means of a proposed internal cognitive device to emulate or 'synthesize' how this would sound when produced. By doing so, they can check the accuracy of their articulation and repair errors with quicker internal cognitive processing times (Roelofs 2005). Moore (2007) proposed the model Predictive Sensorimotor Control and Emulation (PRESENCE) which builds upon the principles of internal emulations and "feedback loops" by claiming that such emulators can be employed not only for linguistic information but also to predict and simulate the emotional state of the speaker. Moore (2007) does not specify whether the objects of perception are gestural or auditory. Neither of these models, however, benefit from the support of observable, neurophysiological evidence and designing experiments to test their predictions remains difficult. Regardless, such models represent the growing consensus of the field which is that speech production and perception are without a doubt two sides of the same coin and linked in various ways. What remains to be clarified is the nature and cross-subject generalizability of this link and exactly how information is shared between modalities.

In conclusion, sufficient neurophysiological evidence is still lacking in order to offer strong support for any of the previously proposed theories regarding the link between speech production and perception. Overall, more than linguistic evidence is necessary to understand such complex cognitive processes that are activated and employed in all aspects of the human experience. The increased number of neurological studies probing these questions echoes the growing consensus in psycholinguistics and speech science research that language is not a unique, closed system but rather a specific manifestation of cognitive and biological processes shared across many dimensions of human behavior. In this regard, speech perception is generally no longer considered an exclusively linguistic task but is rather seen as a *specialized* use of the auditory system (Moore 2007). Future research in speech perception thus requires a more interdisciplinary approach and should consider the interplay of language with other areas of human cognition.

1.2.2 Empirical evidence of link between acoustic cues in production and perception

Much research has been conducted on the link between acoustic cues used in production and perception. A major focus of this research has been on languages undergoing a sound change given that a mismatch at some level is expected, reflecting the intergenerational change in progress. In other words, speakers need to be able to perceive a contrast they no longer produce to ensure accurate communication with all members of their community, particularly with older speakers who may not have acquired the sound change (Coetzee et al. 2018; Pinget et al. 2020; Brunelle et al. 2020; Tạ et al. 2022 and many more). Results are, however, often conflicting and a clear picture remains to be gained regarding the role of these two modalities in sound change.

Coetzee et al. (2018) observed that the relation between cue weights of f0 and VOT, in an ongoing sound change in Afrikaans where f0 is becoming the primary cue for onset stops, varies across speakers. In particular, most younger speakers, who seldom produce different VOT values, paid attention to differences in VOT less than older speakers in perception. However, some young participants showed a relative mismatch between production and perception. In other words, despite almost exclusively utilizing f0 in production, these speakers still made use of the VOT cue in perception. A similar result is seen in the devoicing of /v/ in Dutch, a process nearing completion for many speakers in production, although nearly all individuals are still sensitive to voicing in perception (Pinget et al. 2020). Chru is believed to currently be undergoing a process of transphonologization of secondary cues present in a stop voicing contrast. Younger speakers maintain the contrast largely through vowel height, while older speakers still present optional voicing. In perception, vowel height was the most pertinent cue across the board. However, VOT was still found to play a role even for speakers who no longer produce this contrast (Brunelle et al. 2020). A similar situation was observed for Chrau, where perception cue weights were found to be more flexible than production cue weights to accommodate the variability present in production (Tạ et al. 2022). Additionally, although a weak relationship between cue weights in production and perception was found at the group level for Chrau speakers, no such link was observed at the individual level. For speakers of Seoul Korean, a language undergoing quasi-tonogenesis, Schertz et al. (2015) found little correlation between individual weights of VOT and

f0 in production and perception. Studies such as these highlight a production-perception mismatch that can be justified in a way through the presence of a sound change.

The presence of a production-perception link between cue weights for properties that are not thought to be experiencing any phonological change has also been observed in some studies, while others have failed to find a robust link. Shultz et al. (2012), for example, found no correlation between production and perception of VOT and f0 cues for English bilabial stops, which are known to covary between voiced and voiceless categories. Idemaru et al. (2012) found no significant production-perception link between usage of absolute and relative duration of singleton and geminate consonants in Japanese. Schertz (2014) found no correlation between participants' production and perception cue weights across multiple acoustic cues (f0, onset F1, VOT and closure duration) in either the Spanish or English voicing contrast. Beddor et al. (2018), on the other hand, found that the amount of nasality English speakers use in production was able to predict with relative accuracy the perceptual acuteness to nasality. A significant correlation between production and perception of Mandarin tones has also been reported (Leung & Wang 2020), however this link was only significant when considering F0 slope and not F0 mean (Leung & Wang 2024). Given that F0 slope is thought to be a more critical perceptual cue for Mandarin tones, allowing distinction of all four tones with greater accuracy than relying solely on mean F0, Leung & Wang (2024) argue that a production-perception link may only be available for highly important cues. A similar positive correlation has been observed for spectral moments in English fricatives /s/ and /ʃ/ by Yu (2019) and Lo et al. (2023). However, Lo et al. (2023) only observed a strong link between production and perception in monolinguals and not in highly proficient native bilinguals even when these speakers maintained a significant contrast in production.

Overall, the fact that a robust production-perception link has been observed in nearly as many studies in which it has not been observed emphasizes the complicated, and likely not automatic, nature of this possible connection. It is important to consider the wide range of methodologies used in previous studies investigating possible cross-modality links. While most quantification methods are entirely valid, it is difficult to compare results across studies that utilize different coefficients or variables to assess cue weighting strategies or categorical boundary locations in production and perception (Schertz & Clare 2020). Additionally, the nature

of the acoustic cues tested and the design of the production and perception tasks themselves likely greatly influence the type of link that can be measured and the magnitude of such a link. In sum, although production and perception clearly influence one another in various ways, the results summarized here seem to hint that the exact relationship between these two modalities may be specific to the individual, underscoring the importance of considering both inter- and intraindividual variability in the production-perception link.

1.2.3 Empirical evidence of link between variability in production and perception

In addition to the link between specific phonological categories in production and perception, the amount of intraindividual variability in production has been shown to correlate with both consistency and acuity in perception. More variability in the production of English vowels, measured by means of acoustic data, has been repeatedly shown to be significantly correlated with less accurate vowel perception (e.g., Perkell et al. 2004; Perkell et al. 2008). Franken et al. (2017) investigated how variability in the production of the Dutch vowel pairs /ɛ-ɪ/ and /ɑ-ɔ/ affects perception of this same contrast. They found higher discrimination accuracy for speakers whose vowels were produced with less variability, as evidenced by a greater inter-phoneme distance in formant space, indicative of more distinct categories. McAllister Byun & Tiede (2017) observed that children between the ages of 9 to 14 who gave more consistent responses in a forced-choice perception task between the sounds /ɹ/ and /w/ in North American English, also produced instances of /ɹ/ with more accuracy, i.e. with more rhoticity. Similarly, Chao et al. (2019) found that the categorical boundary between two vowels in perception was highly correlated with an equidistant point between the centroids of two vowel distributions in production. In other words, the perceptual boundary between the two vowels was further away from the vowel with a higher degree of variability in production. This result was argued to show that patterns of production variability can provide important information regarding the organization and mental representation of sounds.

Harper (2021) analyzed the correlation between production variability and perceptual discrimination accuracy for English /ɹ/ and /s/ and found that the strength of this relation

depended greatly on the acoustic dimension measured and the size of the difference among stimuli. A 4IAX perceptual discrimination task was used to test individuals' perception across three acoustic continua: a F3 continuum for /ɹ/, a center of gravity (COG) continuum for /s/ and a kurtosis continuum for /s/. As a reminder, in a 4IAX task, participants hear four acoustic stimuli and are instructed to choose whether the second or third stimulus differs from the rest. For each trial, the acoustic difference between the three *same* stimuli and the one *different* stimulus was set to be either small, intermediate or large. No correlation between production variability and perceptual discrimination accuracy was found for /s/ for any stimulus group. For /ɹ/, a strong correlation between these parameters was only observed for the tokens with the largest, most salient inter-stimulus difference of F3. Harper (2021) noted that the most robust acoustic-articulatory mapping observed was precisely for F3 of /ɹ/. It was argued then that perceptual accuracy is only enhanced for cues with a strong articulatory-acoustic correspondence, partially coinciding with the results from Leung & Wang (2024). Overall, this type of link has been interpreted as showing that more variability in production is an indicator of less distinct cognitive phonological categories and thus, fine-grained perceptual differences are less easily achieved due to the large area in which a specific segment is represented mentally (Harper 2021). Besides the negative result for /s/ found in Harper (2021), one additional recent study looking at the English vowels /ɛ/ and /æ/ likewise found no significant correlation between production and perception variability (Cheng et al. 2021). Specifically, Cheng et al. (2021) found that perceptual consistency in an identification task did not correlate with overall production variability nor with distinctness of vowel categories.

In sum, research on the relationship between variability in production and perception has presented inconsistent results. While various studies find that participants with less variable production are more sensitive to differences in perception, this is not always the case. It seems that the presence of such a relationship may depend on additional factors, such as the salience of the acoustic cue analyzed or the type of perception task itself. It would additionally be pertinent to understand how this link, on an individual level, may differ depending on the precise phonological segment or phonetic cue studied. Much of the research investigating this link has looked at formant values in vowels, and nearly all of the previous studies on correlations between

production and perception variability have focused on English. It would therefore be of interest to investigate similar relations across a varied selection of acoustic cues and in another language, such as Laurentian French.

1.3 Research questions

The current study seeks to address multiple questions related to intraindividual variability and inter-modality correlations in speech production and perception. Three main questions can be established that aim to provide greater clarity on the unresolved issues presented in the previous sections.

1) How does intraindividual variability in acoustic cues pertaining to different contrasts (stop voicing, vowels /e-ɛ/ and /u-o/ and coronal fricatives /s-ʃ/) vary or fail to vary across individuals in speech production and perception?

1.1) Is the magnitude of variability correlated across segments within a single speaker?

1.2) Does the degree of variability found within the speech of an individual differ across experimental sessions?

2) Is there a link (similar cue weights or categorical boundaries) on an individual level between production and perception?

2.1) Is the stability or robustness of this link dependent on the magnitude of intraindividual variability in either production or perception?

2.2) Is the magnitude of this production-perception link correlated across different segments and acoustic cues?

3) Do individuals with higher variability in production also exhibit higher variability (less consistency) in perception?

3.1) Does the presence or magnitude of this correlation depend on the specific segment?

In order to gain an accurate picture of the magnitude and structure of variability in production and perception across various areas of phonetic and phonological space, it is pertinent to consider a wide variety of speech categories. Laurentian French offers various phonological contrasts well-suited for the purposes of this dissertation, namely (1) stop voicing, (2) two vocalic contrasts occupying similar vowel space regions (/e-ɛ/ and /u-o/) and (3) place of articulation of the voiceless coronal fricatives /s/ and /ʃ/.

1.4 Laurentian French

Laurentian French is the principal dialect of French spoken in Canada with over seven million native speakers (Côté 2012). It owes its name to the Saint Lawrence River Valley along which it originated, located in the French-dominant province of Quebec (Côté 2012), and has been spoken in the region corresponding to modern-day Eastern Canada since the 1600s when French colonists settled in North America. Laurentian French is the dominant dialect in Quebec and all other provinces of Canada to the west, thus excluding the Atlantic provinces (such as New Brunswick and Nova Scotia) where Acadian French is the principal dialect spoken (Côté 2012). The situation of Laurentian French is often described as quasi-diglossia due to the important distinction between the colloquial, spoken variety and the standard, written variety, the latter of which is highly similar to the standard variety of Metropolitan French (Chantefort 1976; Auger 2005). However, as in most cases of diglossia, these variants coexist and speakers produce utterances combining aspects of both frequently (Chantefort 1976). In addition, due to the bilingual nature of Canada, many Laurentian French speakers are also fluent in English, whether as highly proficient L2 speakers or as bilinguals from birth with two L1s (Ray & Gilbert 2019). This is especially true of the Ottawa-Gatineau region due to its location on the border between Quebec and the majority-English-speaking province of Ontario. It is noted that the data in this dissertation were collected at the University of Ottawa, a fully French-English bilingual institution located precisely in this area. This pervasive bilingualism has had an impact primarily on the lexical attributes of Laurentian French and has affected the phonetic and phonological system to a much less degree (Auger 2005).

Although in reality, Laurentian French constitutes a group of regional varieties with differing linguistic attributes, various key characteristics remain the same across most varieties and distinguish this variety from standard Metropolitan French (Auger 2005; Côté 2012). Most relevant to the current study are the phonetic differences, of which there are many. Some of the most notable phonetic differences present in Laurentian French but absent in standard Metropolitan French are presented below:

1. The maintenance of contrast between the nasal vowels /ɛ̃/ and /œ̃/ (Côté 2012).
2. The affrication of /t/ and /d/ when preceding the high vowels /i/ and /y/, resulting in allophones of [ts] and [dz] respectively (Dumas 1987; Friesner 2010).
3. The diphthongization of both oral and nasal vowels in closed syllables, such as in the word *père* 'father' /pɛʁ/ which, among other variants, can be pronounced as [paɛʁ]. The presence and salience of these diphthong variants depend on various factors such as regionality and speech register (Côté 2012).
4. The laxing of high vowels in closed syllables resulting in the allophonic changes of /i/ → [ɪ], /y/ → [ʏ] and /u/ → [ʊ] (Dumas & Boulanger 1982; Dumas 1987; Friesner 2010; Côté 2012).
5. The maintenance of both spectral and durational contrast between front and back low vowels /a/ and /ɑ/ distinguishing words such as *patte* /pat/ 'paw` and *pâte* /pɑːt/ 'dough' (Côté 2012).

Finally, an important syntactic difference unique to colloquial registers of Laurentian French is the use of the post-verbal clitic *tu* /ty/ to form questions. For example, the question "are you coming" can be produced as *tu viens-tu?* /ty.vjɛ̃.ty/ where the first *tu* corresponds to the second person singular informal pronoun and the second enclitic *tu* indicates the interrogative tone (Auger 2005).

1.4.1 Stop voicing

Laurentian French has a two-way voicing contrast in obstruents. Previously, some studies had described the voicing contrast in Montreal French as atypical, where phonologically voiced stops were produced with either lead voicing or short lag VOT and phonologically voiceless stops with short lag VOT (Caramazza et al. 1973; Caramazza & Yeni-Komshian 1974). Such a VOT distribution leads to overlapping values which was believed to be the cause of observed misperception in identification tests (Caramazza et al. 1973). However, in nearly all subsequent studies measuring this contrast in Laurentian French, monolingual Laurentian French speakers were found to produce a VOT contrast that is very similar to that of better-studied European French varieties, namely where voiced stops are nearly always produced with lead VOT and voiceless stops with short or mid lag VOT (Jacques 1984; Ryalls et al. 1990; Jacques & Gurlekian 1992; Ryalls et al. 1997; MacLeod 2006; Fowler et al. 2008; MacLeod & Stoel-Gammon 2009; Louër 2019). Sundara (2005), however, observed results similar to the initial studies by Caramazza et al. (1973) where voiced and voiceless stops had overlapping VOT values, although only for dental stops. In addition to VOT, other acoustic cues have been shown to play a role in upholding the voicing distinction in Laurentian French. Jacques (1984) and Jacques & Charbonneau (1989) observed that vowels following phonologically voiceless stops were produced with a longer duration, a higher f0 and less marked F1 transitions. Bilingual English-Laurentian French speakers have been shown to have more varied VOT values than those of monolingual speakers. Specifically, bilinguals have a shorter period of lead voicing in voiced stops and longer duration of aspiration in voiceless stops (Caramazza et al. 1973; MacLeod & Stoel-Gammon 2005; MacLeod 2006; Sundara et al. 2006; Fowler et al. 2008; MacLeod & Stoel-Gammon 2009).

It is well-known cross-linguistically that voicing is difficult to initiate and sustain, due to the precise aerodynamic conditions necessary to produce stable, periodic vocal fold vibrations (Ohala 1983; Solé 2018; Hauser 2021). Devoicing of phonologically voiced stops is a relatively common phenomenon that has been observed in various languages such as English (Dmitrieva et al. 2015; Davidson 2016), German (Pape & Jesus 2011) and European Portuguese (Pape & Jesus

2014), with considerable variability in the frequency and extent of this devoicing. Instances of voicing dying out during the closure before the release have also been observed for Laurentian French, with a higher proportion of devoicing in the speech of female speakers (Schweizer et al. 2025). The observation of such partial closure voicing likely explains the unusual results of Sundara (2005) and Caramazza et al. (1973), where voiced stops produced with incomplete closure voicing were treated as only having positive VOT. Such devoicing tendencies seem to be language-specific as other languages, such as Spanish (Shih et al. 1999; Dmitrieva et al. 2015), Italian (Pape & Jesus 2014) and Hindi (Shih et al. 1999), have been shown to have much lower rates of devoicing. The large amount of inter- and intraindividual variability regarding these instances of devoicing remains relatively unexplored. Due to this, and the fact that previous studies on intraindividual variability have rarely considered the phonetic realization of voicing, voiced and voiceless stops were selected as the main category to be tested in this dissertation.

1.4.2 Vowels

Laurentian French has a rich vowel system, composed of both oral and nasal vowels, with subtle phonemic contrasts between certain vowel pairs, such as /e-ɛ/ and /u-o/ (Martin 2002; Tennant 2012). Cross-linguistically, speakers' production of vowels generally do not correspond to extremely precise acoustic targets but rather to areas of distribution structured within the vowel space unique to each language (Strange et al. 2007). This inherent variability in the production of vowels is especially relevant when considering the need for accurate perception of vowels located close to one another in this vowel space (Brajot et al. 2013; Chao et al. 2019; Cheng et al. 2021, among many others). For the studies in this dissertation, the front vowel /e-ɛ/ contrast and the back vowel /u-o/ contrast were chosen, both of which involve vowels produced close to one another in acoustic space. The /o-ɔ/ contrast would constitute a more direct parallel to the /e-ɛ/ contrast, but Laurentian French phonology does not allow the vowel /ɔ/ in open syllables. As will be explained in more detail in Section 2.1.2, all target words in the current study were in open syllables to ensure the vowels were produced as monophthongs, given vowels in closed syllables can be produced as diphthongs (Côté 2012).

Although few studies provide precise measures of Laurentian French vowels, MacKenzie & Sankoff (2010) measured the first and second formants of six speakers from Montreal and observed consistently distinct categories for the vowel pair /e-ɛ/ and only a slight overlap for the vowel pair /u-o/. Leblanc (2012) observed highly similar results to those reported in MacKenzie & Sankoff (2010) for ten speakers from the cities of Quebec and Chicoutimi. In a more recent study, Riverin-Coutlée et al. (2023) observed consistently distinct categories for the vowel pair /e-ɛ/ in eight speakers from Quebec City. Again, some overlap was observed for /u-o/, although mean formant values remained distinct. Average values of F2-F1 distance from each of these studies is presented in Table 1.1. Although there are considerable differences across studies, the overall magnitude of distinction between the four vowels remains largely comparable.

Table 1.1 F1-F2 distance in Hertz of the four vowels of interest in the current dissertation, reported in three previous studies on Laurentian French.

Vowel	Riverin-Coutlée et al. (2023) *n = 8*	MacKenzie & Sankoff (2010) *n = 12*	Leblanc (2012) *n = 10*
/e/	2231 Hz	1750 Hz	1200 Hz
/ɛ/	1694 Hz	1150 Hz	880 Hz
/o/	385 Hz	500 Hz	272 Hz
/u/	394 Hz	650 Hz	371 Hz

1.4.3 Fricatives

Laurentian French distinguishes two coronal fricatives: dental /s/ and postalveolar /ʃ/ (Walker 1984; Dumas 1987). In the speech of adult Laurentian French speakers, Toda (2009) observed distinct Center of Gravity (COG) values, corresponding to the frequency containing the highest concentration of energy, for /s/ and /ʃ/. Laurentian-French-speaking children also distinguish the fricatives primarily by COG, though differences were also observed for other spectral measures, such as spectral variance, i.e. standard deviation, skewness and kurtosis,

although the size of differences in these secondary spectral cues depended greatly on age (Salinas Marchant 2025). In other languages, various acoustic properties have been found to be pertinent in distinguishing place of articulation in fricatives. In English, for example, in addition to COG, onset F2 frequency and spectral skewness have also been shown to vary in production and play a role in perception (Jongman et al. 2000; Li et al. 2011). Using machine learning, Ulrich et al. (2021) observed that COG and spectral spread alone were sufficient to distinguish Russian /s/ and /ʃ/. However, Ulrich et al. (2023) observed a large degree of interindividual variability in the use of spectral cues, where some speakers, for example, consistently produced stable peak values for each fricative and others produced a much more variable range of peak frequencies. Overall, given the observations across various languages, including Laurentian French, that COG is the most salient acoustic property distinguishing coronal fricatives, it is predicted that this acoustic measure will provide the most reliable information in this study.

1.5 Experimental design

To address the research questions outlined in Section 1.3, acoustic recordings in a production experiment and response data from a two-alternative forced choice (2AFC) perception identification task were collected. Initially, the production data of ten participants was recorded to ascertain pertinent acoustic ranges for the subsequent perception experiment. This initial production experiment is referred to as *Production A*. A few months later (between three and five), these same ten participants returned and participated in a single-session, combined production and perception experiment, along with 48 additional participants. This second production experiment is referred to as *Production B*. In both production experiments, acoustic data was collected from multiple repetitions of the same list of target words embedded in a frame sentence. Immediately following the recording task in Production B, participants carried out the perception task, which consisted of two types of continua and three phonological contrasts. The outline of the three experiments (Production A, Production B and the perception experiment) is shown in Figure 1.1.

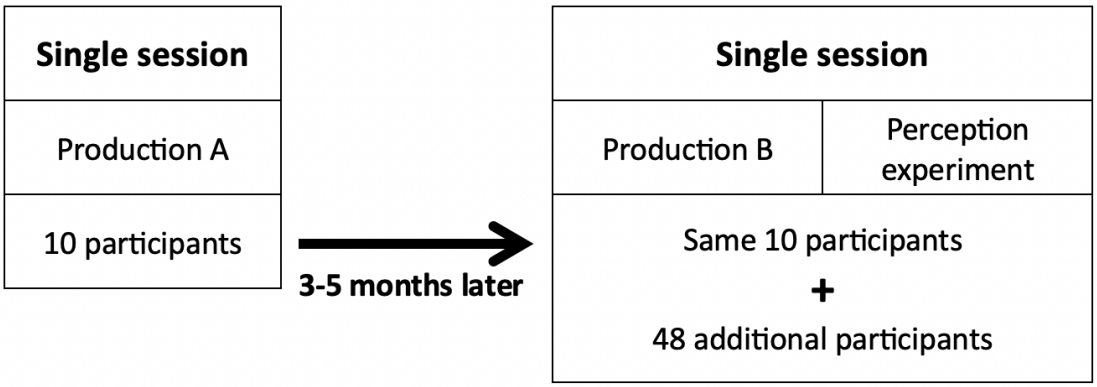


Figure 1.1 Outline of experimental sessions and data collection. A production experiment (A) was carried out with ten participants. Three to five months later, these same participants returned to participate in a single-session, back-to-back production (B) and perception experiment. Production B was identical to Production A. Forty-eight additional participants carried out the single-session Production B and perception experiment. Of these 48 additional participants, nine were excluded, as explained in Section 2.1.1.

In the perception experiment, the first continuum type tested cue weights by covarying voicing proportion and post-release aspiration duration across four voiced and voiceless stop contrasts (/p-bɛ/, /p-bo/, /k-gɛ/, /k-go/). While both of these acoustic measures are often combined into the commonly used VOT term (Lisker & Abramson 1964), they correspond to two distinct articulatory configurations and are both important in upholding the stop voicing contrast in Laurentian French. Further details on the terminology and definitions of the voicing cues will be provided in Section 2.2.1 and Section 3.1.2.1. The second continuum type tested categorical boundary location by manipulating formant and/or spectral information to create general intermediate steps between two phonemes. Three such continua were used to test categorical boundary location of two vowel contrasts (/e-ɛ/ and /u-o/) and the fricative contrast /s-ʃ/. Different continuum types and acoustic measures were chosen to probe distinct aspects of speech perception and investigate how intraindividual variability and the production-perception link may differ as a function of the phonetic information that is perceived.

Indeed, the three acoustic dimensions investigated in this study arguably correspond to distinct levels of auditory-perceptual processing, given that the most salient acoustic cues

thought to uphold these phonological distinctions are located at different frequency ranges of the acoustic signal. Although gender, age and other sociolinguistic factors have an effect on precise values in Hertz, general tendencies are as follows: voicing occupies the lowest frequencies of the acoustic signal (from 70 to 250 Hz; the burst can admittedly be higher), formant values for vowels have a much wider range and extend to higher frequencies (from approximately 300 to 2500 Hz) and the most salient spectral properties of fricatives are thought to be in the upper end of frequencies used in phoneme categorization (from approximately 3000 to 8000 Hz). Given the nonlinear nature of human hearing and the differences in information load found at different frequency ranges (Tuller et al. 2011), the magnitude of variability may depend on the nature of the contrast in question. Moreover, the nature of the acoustic information utilized in the acoustic signal to uphold these distinctions differs. Whereas both closure voicing and post-release aspiration largely rely on durational cues, the vowel and fricative contrasts tested here are distinguished through spectral cues, although duration can often be a context-dependent secondary cue. To accurately address Research question 1, it is therefore important to consider the type of acoustic information produced and perceived when examining the relationship between variability across distinct segments.

In addition to the type of acoustic information relevant to each contrast, the articulatory constraints imposed by each segment may affect the amount of variability observed in production or tolerated in perception. The production of voiced and voiceless stops relies on precise timing of distinct vocal fold configurations. Voiced stops, in particular, must overcome the aerodynamic voicing constraint (AVC) to ensure stable vocal fold vibrations throughout the closure (Ohala 1983; Solé 2018). The AVC designates the fact that increased air pressure in the supraglottal cavity during stop closure diminishes the transglottal airflow necessary to maintain vocal fold vibration. The heightened demands on articulatory precision may result in greater production variability for stop voicing than for vowels and fricatives. Vowels arguably require less precision due to the relative nature of their production, observable in the commonly observed wide spread of formant vowels defining vocalic space (Strange et al. 2007). In other words, tongue position in vowels permits greater flexibility than other articulatory categories due to the lack of direct constriction in the oral cavity. The articulation of coronal fricatives, however, allow

little leeway regarding the location (dental or alveopalatal) or disposition of the tongue tip (apical or laminar). Additionally, fricatives are thought to need a highly specific articulatory setting in order to accurately produce the necessary turbulent airflow (Ohala 1983, 2011; Ohala & Solé 2010). It could therefore be that there is greater variability in vowels than fricatives in production, i.e. acoustic ranges, and perception, i.e. response consistency. The difference in the primary articulator upholding the acoustic distinctions of the stop voicing contrast, on the one hand, and vowel and fricative contrasts, on the other, makes it difficult to predict how the magnitude of intraindividual variability may vary between the two.

Research question 2 and Research question 3 probe different aspects of a possible production-perception link, that is, similarity in acoustic cue ranges and similarity in the degree of variability of acoustic cues across modalities. As reviewed previously, f0 and formant transitions have been shown to covary with voicing cross-linguistically, including in Laurentian French (Jacques 1984, Jacques & Charbonneau 1989). These cues have been shown to be available in perception, although they are generally thought to be secondary to actual vocal fold vibrations. Given this and the previous observation of some Laurentian French voiced stops being produced with short lag VOT (Caramazza et al. 1973; Caramazza & Yeni-Komshian 1974; Sundara 2005) and the high degree of partial devoicing of voiced stops (Schweizer et al. 2025), it is postulated that some speakers may learn to rely less on actual voicing in perception. If a production-perception link exists at the individual level, such partial-voicers would then weigh voicing less in perception, even when secondary cues they may normally resort to are ambiguous. For vowels and fricatives, a stronger production-perception link would indicate that listeners normalize the perceived sound to align with their own productions and then carry out categorization based on their own phonological boundaries. Due to the lack of studies investigating the possible effect intraindividual variability may have on the production-perception link, it is difficult to make solid predictions regarding Research question 3.

1.6 Outline of dissertation

Chapter 2 describes the experimental design and data analysis of the production experiment, from both Production A and Production B. Chapter 3 presents the experimental design and data analysis of the perception experiment. Inter- and intraindividual patterns concerning the production-perception link and its relationship to variability are explored in Chapter 4. Finally, Chapter 5 discusses the results in light of previous research and compares the observations of this dissertation with similar patterns from various disciplines in order to better understand the function and importance of variability as a central aspect of human behavior. Chapter 6 summarizes the key findings from this dissertation and describes how these results can inform future phonetic studies.

2. Production experiment

A production experiment was carried out to analyze the variability in the three categories tested, namely stop voicing, vowels and voiceless coronal fricatives. In order to address Research question 1.2 regarding the stability of an individual's acoustic cues and variability across experimental sessions, the production experiment was carried out twice. An initial production experiment was conducted with ten individuals, henceforth referred to as *Production A*. The acoustic results of this group were used to inform the acoustic ranges chosen in the stimuli manipulation of the perception experiment, be presented in Chapter 3. After several months, these same participants returned to carry out the combined single-session production and perception experiment. This second round of the production experiment was labeled *Production B*. Production B was carried out in the same experimental session as the perception experiment to ensure the observations in both modalities were directly comparable and representative of a similar timepoint for each speaker. An additional 48 subjects, who did not participate in Production A, participated in this single-session production and perception experiment. This chapter will first discuss the results from Production B to permit an analysis of the general tendencies observed across all subjects. In addition, it is data from Production B that will be subsequently compared with the results from the perception experiment in Chapter 4. The analysis of Production B data will be done in Sections 2.1-2.3. Then, in Section 2.4, the differences and similarities in acoustic values across experimental sessions from the 10 participants who completed both Production A and Production B will be presented.

2.1 Methodology

2.1.1 Participants

Data was collected from 58 participants. No participant reported known auditory or linguistic impediments and all received either course credit or $25 CAD as compensation. Nine participants were removed from the final dataset either due to not being native speakers of French or due to speaking a dialect other than Laurentian French. Therefore, the data of a total

of 49 participants were analyzed. Of these 49 participants, 35 were female, 13 were male and one was non-binary. Their mean age was 21 years old (range: 18 – 45). As noted above, ten of these participants had completed the initial production experiment, Production A, before returning a few months later to carry out the combined production-perception experiment. Due to participants' limited availability, the average time elapsed between Production A and Production B was between 3 to 5 months (mean = 4.5).

To determine the minimum sample size needed to observe analyzable tendencies, a power analysis was carried out using the *pwr* package in R (Cohen 1988; Champely et al. 2017). The variables were set as follows: a significance level of 0.05, a power of 0.8 (Cohen's d indicative of a large effect) and an effect size (Spearman's rho correlation coefficient) of the mean of significant correlations found in previous studies probing intraindividual variation, namely 0.45 (Whalen et al. 2018; Chao et al. 2019; Harper 2021). This resulted in an ideal sample size of 35 to achieve sufficient statistical power. Even though the final count of 49 participants exceeds this threshold, it is important to recognize the limitations of drawing exceedingly strong conclusions from a relatively limited population. Nevertheless, it is believed that the sample size in this study allows for significant and reliable tendencies to be observed that can then be further tested in subsequent experiments.

All participants were native speakers of Laurentian French, originating from and having lived nearly their entire lives in either Quebec or Ontario, and were currently living in Ottawa at the time of data collection. In addition to Laurentian French, all participants spoke at least one additional language, either as native speakers (therefore self-reporting as fully bilingual) or as highly proficient L2 speakers. Table 2.1 presents the distribution of additional languages spoken by all participants.

Table 2.1 Counts of native languages, other than Laurentian French, and L2s spoken by the 49 participants whose data was analyzed.

	Native language	Proficient L2	Intermediate L2
English	13	26	10
Arabic	1		3
Swahili	1		
Serbian	1		
Spanish			9
Italian		1	1
Portuguese			1

2.1.2 Stimuli

To obtain sufficient repetitions of the relevant phonological categories without increasing the total length of the experiment, acoustic measurements were taken from multiple segments within each target word. Specifically, stimuli were disyllabic words with voiced stops (/b, d, g/), voiceless stops (/p, t, k/) and coronal fricatives (/s, ʃ/) in word-medial position followed by one of the four vowels of interest (/e, ɛ, o, u/). Words containing word-medial dental and palatal nasal stops (/n, ɲ/) were also included with the aim of increasing the number of repetitions of each vowel. The preceding vowel was held constant as /a/ with the exception of the words *gagner* and *gagnait*, produced with /ɑ/ in Laurentian French. These two words were chosen given that no real word contains the sequence /aɲe/ or /aɲɛ/. Real words were prioritized to increase the probability that participants produced natural speech. However, two nonce words were included (*dassous* and *dagnou*), as no real word in Laurentian French contains the desired sequences. All target vowels were in open syllables to ensure speakers produced monophthongs, as vowels in closed syllables often undergo diphthongization in Laurentian French (Côté 2012). This resulted in a total of 40 target words as shown in Table 2.2. Given that the largest category of interest in this study concerns stop voicing, the 40 target words were designed to also contain the three

pairs of voiced and voiceless stops in word-initial position, when possible, thus increasing the number of repetitions of these phonemes. Tokens where acoustic measures were taken for word-initial stops as well are marked with grey shading in Table 2.2. Finally, to achieve a total of at least three word-initial tokens for each voiced and voiceless stops, four additional words were included, given in Table 2.3, thus resulting in a total of 24 additional stop tokens. The word-medial segments in the words in Table 2.3 were not analyzed, but the word-final vowels were included in the analysis.

All words were embedded in the carrier phrase *T'entends-tu* _____ *pour vrai?* /tɑ̃tɑ̃-ty ____ puʀ vʀɛ/ 'Do you hear ____ for real?'. The presence of the interrogative clitic *tu*, representative of an informal language register, was chosen to encourage participants to produce more natural language rather than formal, careful speech. Tokens were placed utterance-medially to encourage speakers to produce the target words without a pause, thus allowing for accurate measurement of the onset of stop closure. Target sentences were randomized and presented five times in the same randomized order to participants. Three participants produced only four repetitions due to time constraints. The final ideal count of analyzed segments per subject, before exclusion of tokens due to mispronunciation or recording errors, was 500. Table 2.4 provides a breakdown of this count across the segments of interest.

Table 2.2 Stimuli produced by participants in the production experiment. All second-syllable onsets and word-final vowels are target segments. Grey-shaded boxes represent words for which acoustic measurements were also taken for the word-initial stop.

	/e/	**/ɛ/**	**/o/**	**/u/**
/b/	*abbé* /abe/ 'abbey'	*rabais* /ʀabɛ/ 'discount'	*labo* /labo/ 'lab'	*tabou* /tabu/ 'taboo'
/p/	*taper* /tape/ 'to press'	*clapet* /klapɛ/ 'valve'	*drapeau* /dʀapo/ 'flag'	*papou* /papu/ 'Papuan'
/d/	*gradé* /gʀade/ 'rank'	*cadet* /kadɛ/ 'cadet'	*cadeau* /kado/ 'gift'	*gadoue* /gadu/ 'mud'
/t/	*daté* /date/ 'dated'	*ratait* /ʀatɛ/ 'she/he/it was missing'	*bateau* /bato/ 'boat'	*atout* /atu/ 'skill'
/g/	*blaguer* /blage/ 'to joke'	*pagaie* /pagɛ/ 'oar'	*magot* /mago/ 'loot'	*ragoût* /ʀagu/ 'stew'
/k/	*laquer* /lake/ 'to lacquer'	*paquet* /pakɛ/ 'package'	*tacot* /tako/ 'beat-up car'	*à coup* /aku/ 'fits and starts'
/s/	*assez* /ase/ 'enough'	*traçait* /tʀasɛ/ 'she/he/it was tracing'	*l'asso* /laso/ 'the association' (abbreviation)	*dassous* /dasu/ *nonce word*
/ʃ/	*cacher* /kaʃe/ 'to hide'	*sachet* /saʃɛ/ 'small bag'	*cachot* /kaʃo/ 'dungeon'	*cachou* /kaʃu/ 'cashew'
/n/	*planer* /plane/ 'to glide'	*panait* /panɛ/ 'she/he/it was breading'	*panneau* /pano/ 'sign'	*à nous* /anu/ 'to us'
/ɲ/	*gagner* /gɑɲe/ 'to win'	*gagnait* /gɑɲɛ/ 'she/he/it was winning'	*agneau* /aɲo/ 'lamb'	*dagnou* /daɲu/ *nonce word*

Table 2.3 Additional stimuli in the production experiment where acoustic measurements were taken for the word-initial stops and the word-final vowels.

/b/	/d/	/g/
balai /balɛ/ ‘broom’	*daller* /dale/ ‘to tile’	*gager* /gaʒe/ ‘to bet’
bazou /bazu/ ‘beat-up car’		

Table 2.4 Number of tokens produced by each participant for each segment of interest. Three participants produced four repetitions instead of five and thus have slightly lower counts.

Total token count							
Voiced stops		**Voiceless stops**		**Vowels**		**Fricatives**	
/b/	35	**/p/**	45	**/e/**	60	**/s/**	20
/d/	40	**/t/**	35	**/ɛ/**	55	**/ʃ/**	20
/g/	40	**/k/**	45	**/o/**	50		
				/u/	55		

2.1.3 Procedure

At the start of each experimental session, participants signed a consent form and filled out a language background questionnaire. The questionnaire contained the following questions:

1. Name
2. Gender
3. Age
4. Place of residence and previous places of residence (with duration)
5. Native languages
6. Non-native languages and proficiencies
7. Percentage spent speaking each language with family, friends and at the workplace/university

Afterwards, participants were shown the list of target words and informed that some were nonce words that should be pronounced according to French orthography. Participants were then seated in a soundproof booth in front of a computer screen. A Shure KSM44a microphone was placed approximately 15cm from their mouth on the side. The acoustic signal was collected with the program AAA (Articulate Assistant Advanced, Articulate Instruments Ltd., 2012). Participants were instructed to speak as naturally as possible and produced multiple test sentences to ensure the microphone was positioned correctly. The researcher then sat outside the soundproof booth and conducted the experiment. In total, the recording session lasted between 15 and 20 minutes. Immediately after this recording session, the participants completed the perception experiment (see Chapter 3).

2.1.4 Analysis of acoustic cues

Acoustic data was analyzed in Praat (Boersma & Weenink 2022). Annotations were realized manually and included the preceding /a-ɑ/ vowel, stop closure and subsequent aspiration of both word-initial and word-medial stops, frication portion of the fricatives and word-final vowels. For tokens where the word-initial stop was also included, the preceding /y/ vowel was also annotated. The precise realization of voicing was measured by placing landmarks at voicing onset and, if necessary, voicing offset and voicing resumption. Figure 2.1 shows an example annotation of the target word *daté* /date/ 'dated' where voicing dies out before the release. The post-release aspiration included both the burst and any subsequent aspiration before the onset of voicing of the following vowel.

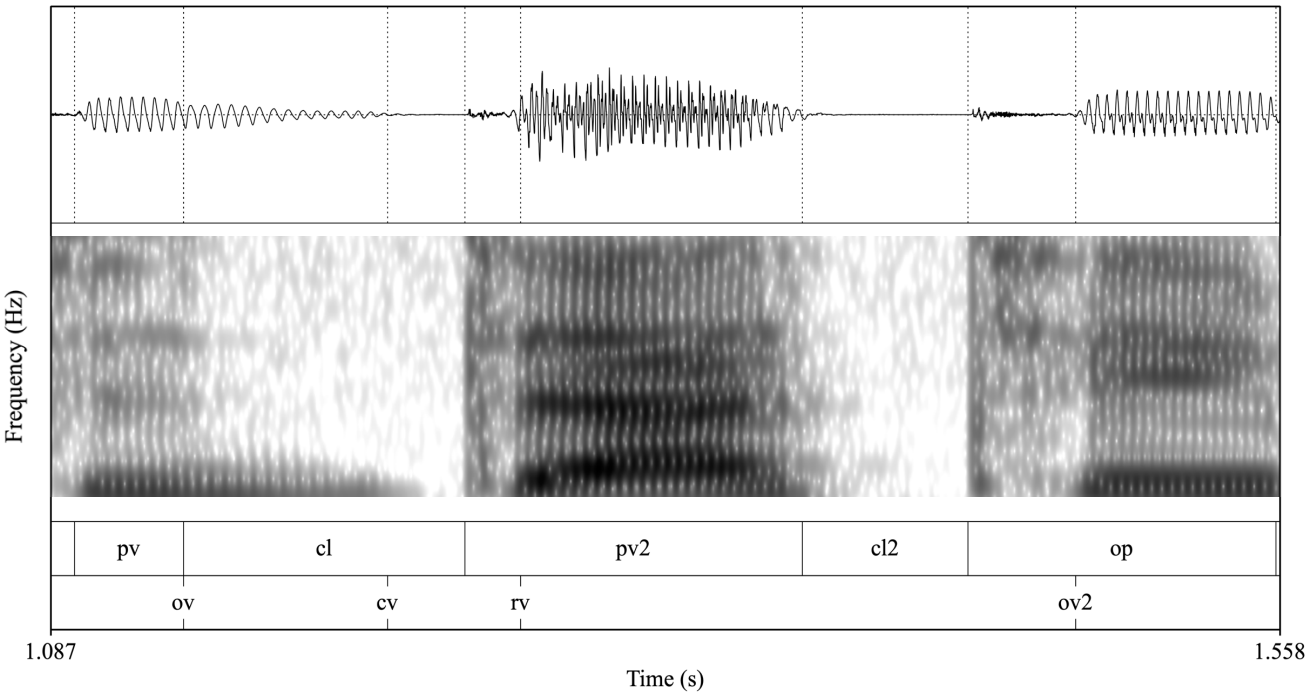


Figure 2.1 Example annotation of target word *daté* /date/ 'dated' with the following annotation labels: *pv* is the previous vowel /y/, *cl* is the closure of /d/, *pv2* is the vowel /a/, *cl2* is the closure of /t/, *op* is the final vowel /e/, *ov* is the onset of voicing of /d/, *cv* is the cessation of voicing of /d/ during stop closure, *rv* is the resumption of voicing of /d/ during the post-release /a/ and *ov2* is the onset of voicing of the final vowel /e/.

Acoustic measurements for stops and vowels were extracted using PraatSauce (Kirby 2018), a Praat implementation of VoiceSauce (Shue et al. 2011). For stops, measurements included the closure duration, voicing duration and post-release aspiration duration, corresponding to the traditional definition of positive VOT (see Section 2.2.1.2 for an explanation of the decision to use this terminology). Word-initial segments produced with a discernible pause before the target word were excluded from analysis, given that it is impossible to accurately identify the start of the closure, but the word-medial segments of these words were analyzed. For vowels, F1 and F2 at every millisecond and total duration were extracted. Removal of formant-tracking errors is explained in Section 2.2. For fricatives, measurements were extracted by means of a Praat script (DiCanio 2013) and included the first four spectral moments averaged

across the center 80% of frication (thus excluding the transitions to and from the preceding and subsequent vowel, respectively). The first four spectral moments correspond to center of gravity, standard deviation, skewness and kurtosis.

2.1.5 Analysis of variability

Intraindividual variability was operationalized by means of the Coefficient of Variation (CoV). CoV measures the dispersion of a set of datapoints as the ratio of the standard deviation to the mean (Everitt 1998), as shown below.

$$CoV = \frac{SD}{mean}$$

It is a standardized, unitless measure which allows for the comparison of properties measured on different scales or with different units and has been used in various studies assessing intraindividual variability in speech (Preston & Koenig 2011; Lee & Sidtis 2017; Harper 2020; Hauser 2022). A higher CoV indicates a greater amount of intraindividual variability. A CoV of zero indicates that all datapoints are identical. In addition to its use in phonetics, the CoV is also commonly used to quantify intraindividual variability across a wide range of fields, such as immunology (Reed et al. 2002), plant biology (Jo et al. 2024; Trevizan et al. 2026) and motor sciences (Getchell et al. 2001; Haar et al. 2017). Its use in a wide range of studies across distinct types of data demonstrates its utility as a metric for relative variability.

While intraindividual variability can be quantified by absolute measures, such as the standard deviation (SD), the absolute range or the interquartile range, the CoV is more appropriate in this study for two main reasons. First, given that it is a unitless measure, it permits direct comparison across phonological categories and acoustic measures, which is crucial to the research questions of this dissertation. Second, absolute SD cannot be fully interpreted without comparison to an individual's baseline values. Indeed, the same degree of dispersion is indicative of distinct degrees of variability for individuals or categories with different mean values. For

example, the magnitude of variability indicated by a SD of 10 ms of post-release aspiration is different for an individual with a mean of 20 ms (CoV of 0.5, thus greater variability) and an individual with a mean of 50 ms (CoV of 0.2, thus lower variability). Given that the focus of this dissertation is *relative* rather than absolute intraindividual variability, the CoV is a well-suited metric.

A potential methodological concern is that an increase in the mean will naturally decrease the resulting CoV, if the SD remains constant. While this is indeed a mathematical consequence of the formula, it does not represent an inherent flaw in the measure, but simply confirms the utility of CoV in measuring relative variability. Given the nature of the CoV, if a given phonological category is found to have a higher mean but a lower SD, this is indicative of a real behavioral phenomenon, whereby the individual may naturally aim to control their acoustic output for that category with greater precision. That being said, this potential confounding factor will be taking into consideration and interpretations of the variability effects will be treated with the necessary caution.

Given that linguistic communication requires speakers to avoid wild fluctuations, CoV values in the type of acoustic data analyzed in this dissertation are not expected to exceed a value of 1, which would indicate a very large degree of variability, where the standard deviation is equal to the mean. A separate CoV was calculated for each individual's productions of a specific acoustic measure.

2.1.6 Statistical analysis

To better understand intraindividual variability, cross-speaker tendencies also need to be established. Given this, two main types of statistical analyses were undertaken. For aggregated data, where speakers' productions were pooled together, linear mixed models were carried out. The general approach taken was to start from the maximal model (including relevant fixed effects and all interactions) with by-subject random intercepts and random slopes when relevant. Fixed effects (with reference levels bolded) were as follows: VOICING (**voiced**, voiceless), PLACE (**labial**, coronal, velar), SEGMENT POSITION (**word-initial**, word-medial), word-medial VOWEL (**/e/**, /ɛ/,

/o/, /u/) and FRICATIVE (**/s/**, /ʃ/). Random slopes were only included if they allowed the model to converge. Non-significant main effects and interactions were removed if they did not improve model fit, starting with the least significant element, assessed by comparing models and choosing the one with a lower AIC. These models were carried out in R (R Core Team 2022) using the *lmerTest* package (Kuznetsova et al. 2017). Details regarding the structure of specific models will be presented throughout the text in this section. When significant interactions involved multiple levels of multiple predictors, these were further analyzed with post-hoc Tukey tests using the *emmeans* package (Lenth & Piaskwoski 2022). Intraindividual variability across segments was analyzed by means of Pearson correlations of the Coefficient of Variation. Running multiple correlation tests can increase the risk of false positives, where significant correlation results are not necessarily indicative of true relationships. To account for this, the significance threshold is often lowered. Given that for most of the categories tested here, the number of correlations rarely exceeded five, the significance threshold in this dissertation was maintained at $p < .05$, but strong significance is interpreted as being $p < .01$.

It should be noted that the experimental design presents a limitation with regard to the analysis of Segment position. Given that the vowel preceding word-medial consonants was held constant as /a/, all word-initial stops precede /a/. Word-medial stops, on the other hand, were followed by one of the four vowels of interest in this study: /e, ɛ, o, u/. This means stops in word-medial and word-initial contexts always precede different vowels. To ensure that differences between segment positions can be attributed to positional differences rather than following vowel effects, closure duration of stops in both positions was analyzed (see Figure 2.2). Indeed, stops in word-initial position were produced with longer closure durations, a known characteristic of prosodic boundaries in various languages, including European French (Fougeron 2001; Keating et al. 2003). This difference in closure duration is highly significant ($\beta = -18.09$, $t = 7.42$, $p < .001$, see Table A1). It can therefore be assumed that the differences observed between segment positions are mostly due to distinct prosodic boundaries and not artefacts of following vowel identity.

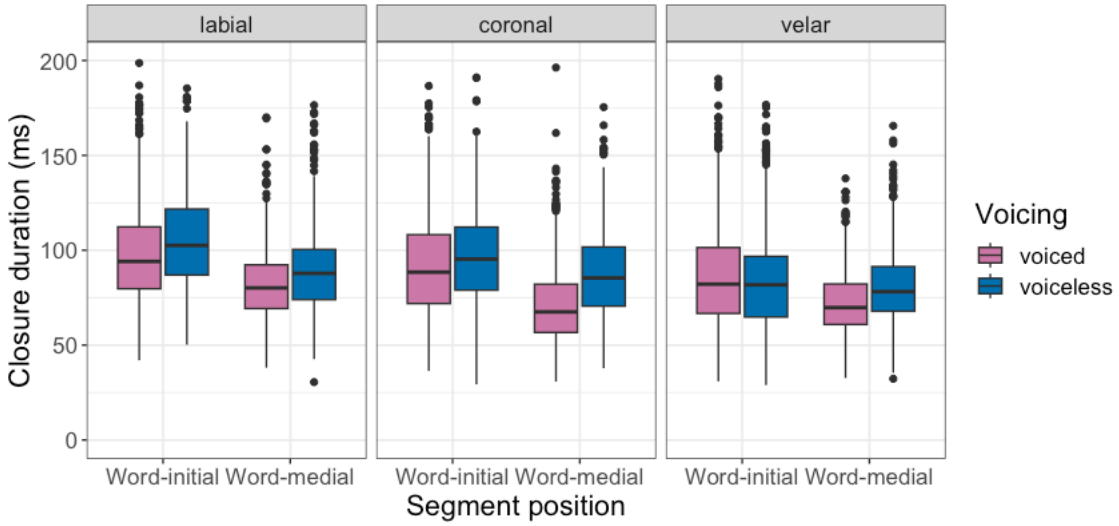


Figure 2.2 Closure duration of voiced and voiceless stops by segment position and place of articulation.

2.2 Results

2.2.1 Voiced and voiceless stops

Voiced and voiceless stops were produced with a large amount of both inter- and intraindividual variability, specifically regarding the realization of closure voicing. That is, many voiceless stops were produced with short durations of voicing at the start of the closure and many voiced stops were produced with incomplete voicing followed by some degree of post-release aspiration. Such inter- and intraindividual variability in the realization of voicing was previously reported by Davidson (2016). Incomplete closure voicing for voiced stops in so-called "true-voicing languages" has also been reported in Laurentian French (Schweizer et al. 2025) and European Portuguese (Pape & Jesus 2014). Figure 2.3 presents the five main patterns of stop voicing observed in this study: (A) complete closure voicing with no post-release aspiration, (B) complete closure voicing with post-release aspiration, (C) >50% closure voicing with post-release aspiration, (D) < 50% closure voicing with post-release aspiration and (E) voiceless closure with post-release aspiration. Pattern (A) corresponds to a canonical voiced stop while pattern (E) represents a canonical voiceless stop. Henceforth, pattern (C) is labeled as *wane voicing* and pattern (D) as *carryover voicing*.

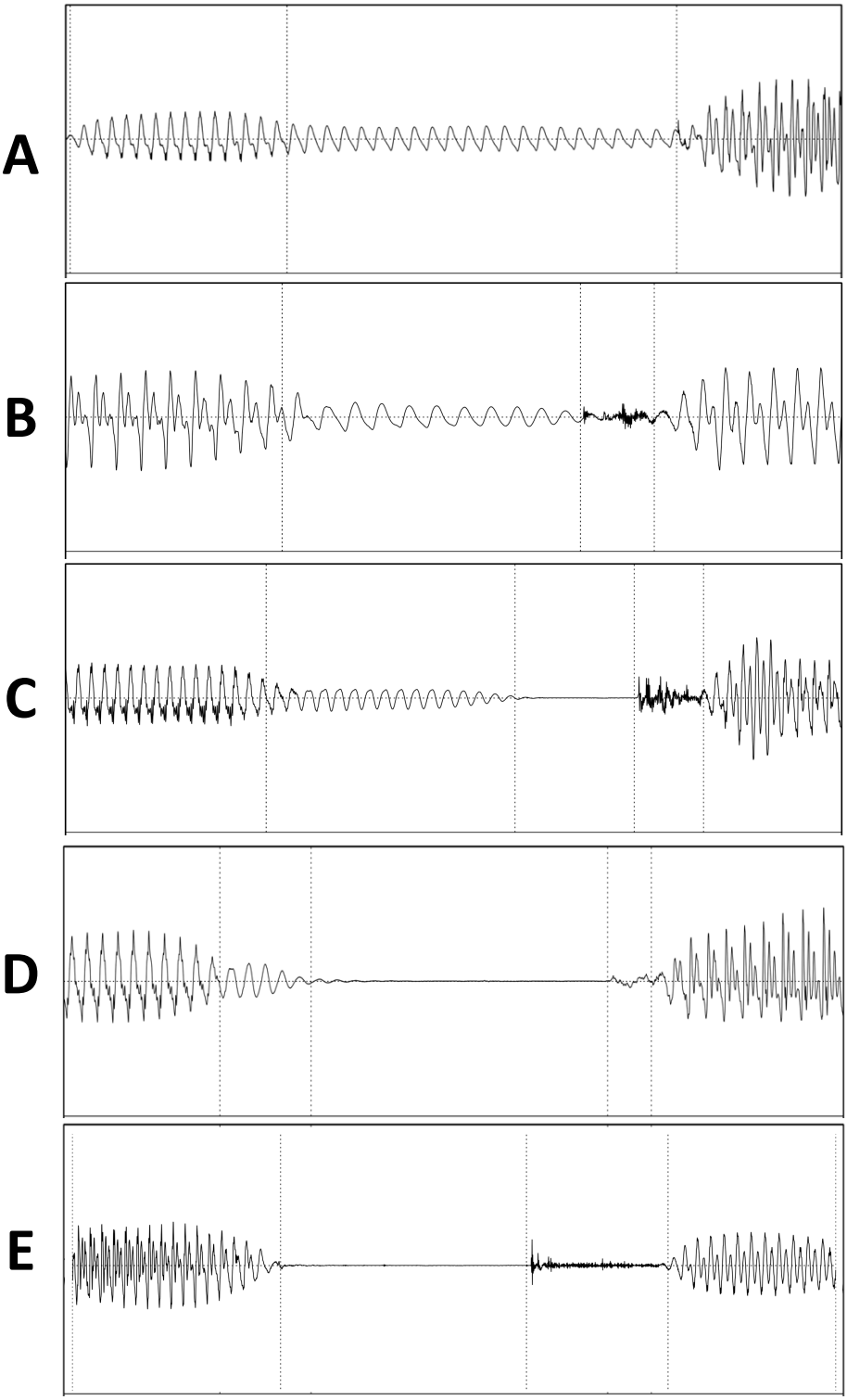


Figure 2.3 Waveforms of the five observed stop voicing patterns: (A) complete closure voicing with no post-release aspiration, (B) complete closure voicing with post-release aspiration, (C) wane (>50%) closure voicing with post-release aspiration, (D) carryover (<50%) closure voicing with post-release aspiration, (E) voiceless closure with post-release aspiration.

According to Davidson (2016)'s definitions of partial voicing based on English data, categories (C) and (D) would both be considered examples of "bleeding". It is important to make a finer-grained distinction in the current study given the two distinct behaviors observed, namely a "dying-out" or wane of voicing for voiced stops, as in (C), and carryover voicing from the previous vowel for voiceless stops, as in (D). These two patterns are arguably the result of distinct articulatory processes and are conditioned by distinct phonological categories. Whereas carryover voicing is generally due to sustained vocal fold vibrations of the previous vowel during the initial portion of the voiceless stop closure (due to a possible misalignment of timing gestures), wane voicing is likely due to a lack of transglottal airflow as a result of an equalization of supra- and subglottal pressure.

It should be noted, however, that the proportion of voicing during closure is influenced by both closure duration and voicing duration. In other words, stops with identical voicing durations could have considerably different voicing proportions if their closure durations are equally as different. Theoretically, this could indicate that some instances of stops labeled as "carryover" could have been produced with the same articulatory process as a stop produced with wane voicing, if the closure duration of the former was much longer than that of the latter. While it is important to consider the interplay of these three variables (voicing duration, closure duration and voicing proportion), the observation that nearly all instances of carryover voicing were found in phonologically voiceless stops and nearly all instances of wane voicing were found in phonologically voiced stops, strengthens the argument that these two patterns stem from distinct articulatory processes.

It is noteworthy that two possible patterns were completely absent in the data: partially voiced stops with no aspiration and voiced stops with initiation of voicing after the start of the closure and then continuous voicing into the following vowel, as presented in the model waveforms in Figure 2.4. The lack of partially voiced stops with no aspiration is relatively unsurprising given that the voicing of post-release vowels following a voiceless closure almost always involves a (most often very short) transition period of voicelessness, encompassing the burst and subsequent aspiration. The lack of voiced stops where voicing did not immediately begin at closure onset is more intriguing, given that in many studies on voiced stops, this

realization is taken as the most common. Indeed, voiced stops are often classified as having a *negative VOT*, which is defined as the duration of voicing taking closure offset as the starting point (Lisker & Abramson 1964). Given that all stops in the current study were in intervocalic position, where the voicing of the previous vowel naturally flows into the subsequent closure, this commonly reported pattern of (largely utterance-initial) voiced stops was articulatorily highly improbable. This pattern of voicing was indeed observed for some voiced stops produced after a significant pause. However, such instances of voiced stops are no longer in intervocalic but rather utterance-initial position. As mentioned previously, since it is impossible to determine the start of the closure, and therefore obtain an accurate measure of closure duration, in utterance-initial stops, these were not included in the final analysis.

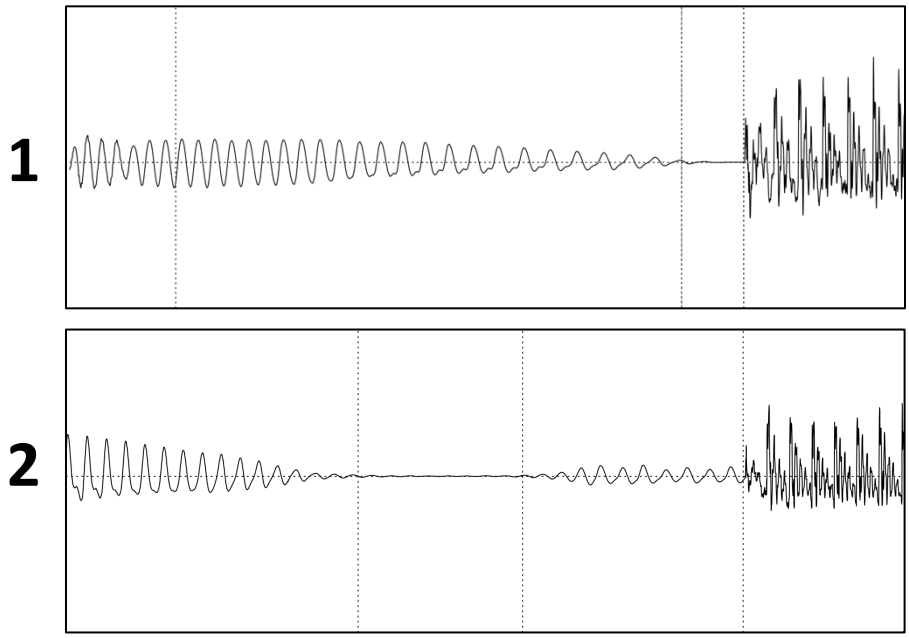


Figure 2.4 Model waveforms of two stop voicing patterns that were not observed in the production data: (1) partial closure voicing that ceases before release without post-release aspiration, (2) partial closure voicing that begins at closure midpoint and continues until the release.

The original definition of Voice Onset Time (VOT) divides stops into two categories: positive, measuring post-release aspiration, and negative, measuring lead-voicing duration taken

from the onset of voicing up until the release of the closure (Lisker & Abramson 1964). However, this categorization misrepresents stops produced with both partial closure voicing that disappears before closure release and subsequent aspiration. For example, based on this definition, the voicing patterns presented in (C) and (D) would be classified as having positive VOT, disregarding the voicing during closure. For that reason, voicing and aspiration are treated as separate phenomena in this dissertation and thus analyzed independently. To avoid the terminological ambiguity of the term VOT, henceforth the terms *voicing proportion* and *post-release aspiration (duration)* will be used. Voicing proportion refers to the percentage of the closure realized with vocal fold vibrations. Post-release aspiration refers exclusively to the delay between the stop release and voicing onset in the following vowel. This division allows patterns of wane and carryover voicing in both voiced and voiceless stops to be analyzed independently of what occurs after closure release. Additionally, the expected aspiration of voiceless stops can be directly compared with the observed aspiration of both partially and fully voiced stops. Such a distinction has proven crucial in understanding the finer details of acoustic realizations of various phonological contrasts across a wide range of languages, such as Armenian (Seyfarth & Garellek 2018), Lebanese Arabic (Al-Tamimi & Khattab 2018), Japanese (Tanner et al. 2020), Raglai (Brunelle et al. 2022), Jarai (Brunelle et al. 2024), Mnong Râlâm (Brunelle et al. 2025), Javanese (Brunelle et al. 2025) and Seoul Korean (Schweizer et al. 2026).

2.2.1.1 Proportion of voicing

Overall, 44% of voiced tokens were realized with incomplete closure voicing (n=1506/3442) and 50% of voiceless tokens were realized with carryover voicing (n=1787/3601). Figure 2.5 presents histograms of voicing proportion for voiced and voiceless stops across places of articulation and segment position. Despite a large degree of interindividual variability, there is nearly no overlap between the voiced and voiceless categories.

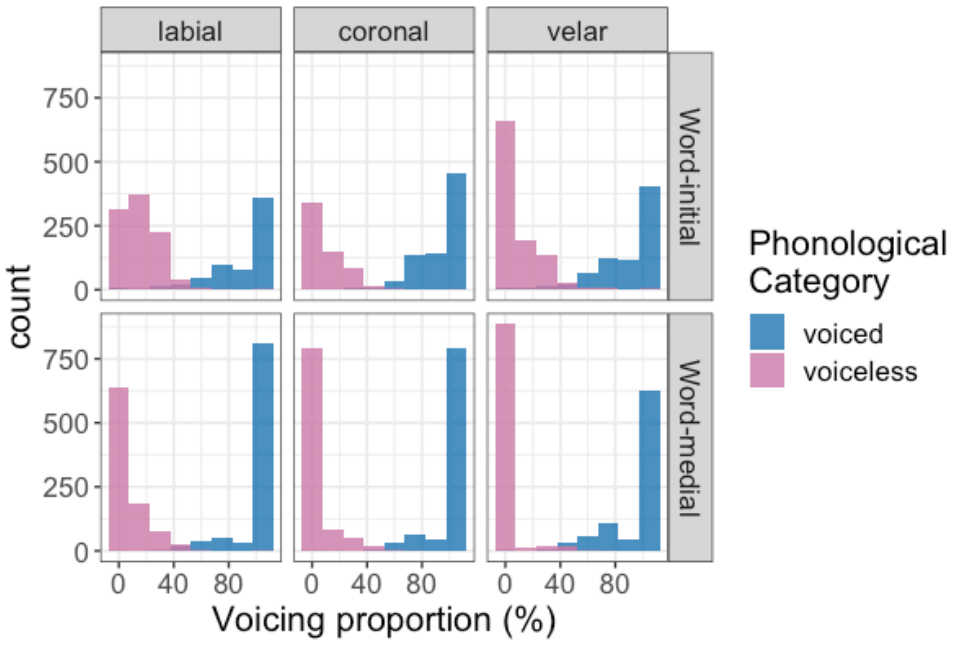


Figure 2.5 Counts of voiced and voiceless stops produced with different voicing proportions in closure.

Two mixed linear regression models were fitted with the dependent variable of voicing proportion, one to the subset of data containing voiced stops and one to the subset of data containing voiceless stops. Fixed effects were Place, Segment Position and their interaction with by-word random intercept and by-subject random intercept and slope for Segment Position. The model results are shown in Table A2 for voiced and Table A3 for voiceless.

Overall, voiced stops were produced with significantly higher voicing proportions in word-medial position ($\beta = 8.64$, $t = 4.7$, $p < .001$). This is likely due to the overall shorter closure durations presented in Figure 2.2. Regarding place effects, post-hoc analyses revealed that word-initial coronal stops were produced with greater voicing proportion than both labial ($\beta = 3.9$, $t = 3.03$, $p = .007$) and velar stops ($\beta = 4.9$, $t = 3.9$, $p = .002$), whereas no difference was observed between word-initial labial and velar stops ($\beta = 0.87$, $t = 0.66$, $p = .8$). Word-medially, however, both coronal ($\beta = 4.7$, $t = -0.9$, $p < .001$) and labial stops ($\beta = 4.1$, $t = 3.4$, $p = .002$) were produced with greater voicing proportion than velar stops, but no difference was observed between coronal and labial stops ($\beta = 0.6$, $t = 0.5$ $p = .9$). Voiceless stops were overall produced with significantly less carryover voicing in word-medial position ($\beta = 8.2$, $t = 5.5$, $p < .001$). Post-hoc

analyses showed that in both word-initial and word-medial positions, voiceless labial stops were produced with greater voicing proportions than both coronal (**Word-initial:** β = 5.6, t = 4.5, p < .001; **Word-medial:** β = -2.9, t = -2.4, p = .04) and velar stops (**Word-initial:** β = 5.9, t = 5.4, p < .001; **Word-medial:** β = 5.3, t = 4.4, p < .001). No differences were observed between voiceless coronal and velar stops in either position (p > 0.1). For both voiced and voiceless stops, velar stops are consistently produced with lower rates of closure voicing in both contexts. This is likely due to the Aerodynamic Voicing Constraint (Ohala 1983; 2011). According to the AVC, the greater difficulty in maintaining vocal fold vibrations at more posterior places of articulation is due to the reduced volume of the oral cavity which shortens the time until pressure equilibrium is reached, thus resulting in the cessation of transglottal airflow.

To determine what effect the following vowel may have on voicing proportion, two mixed linear regression models on the subset of voiced and voiceless stops exclusively in word-medial position were built with the fixed effects of Place, Vowel and their interaction with a by-subject random intercept (Table A4 and Table A5). By-word random intercept was not included as this corresponded to the interaction Place*Vowel. For voiced stops, no significant differences between vowels were observed (p > .07). For voiceless stops, a post-hoc analysis showed that voicing proportion of /po/ was significantly longer than all other vowels (**/e/:** β = 6.07, t = 9.2, p < .001; **/ɛ/:** β = 8.05, t = 12.2, p < .001, **/u/:** β = 7.25, t = 10.9, p < .001). As this corresponds to a single word (*drapeau*), it is likely an idiosyncratic result specific to this word. Overall, voicing proportion tendencies for both voiced and voiceless stops are affected by segment position and place of articulation but not by the following vowel.

2.2.1.2 Post-release aspiration duration

The distribution of post-release aspiration duration for voiceless stops across all speakers is presented in Figure 2.6. Results coincide with previously reported results for French (Rochet & Nearey 1993), as well as cross-linguistic tendencies, namely that longer durations of aspiration are found in more posterior places of articulation and high vowels (Cho & Ladefoged 1999). This is confirmed by a linear regression with the interaction of Place and Vowel and a by-subject

random intercept. The inclusion of the term Vowel renders the term Segment Position redundant, given that the vowel following word-initial stops was always /a/. Post-release aspiration duration increases as the point of constriction recedes in the oral cavity (**PlaceCoronal:** β = 11.88, t = 11.72, p < .001; **PlaceVelar:** β = 34.07, t = 33.57, p < .001). This tendency was observed for all vowels except /o/ where a post-hoc analysis revealed no significant difference between /to/ and /po/ (β = 1.78, t = 1.75 p = .2). Additionally, the back vowels /u/ and /o/ condition longer aspiration than the other vowels (**Vowel/u/:** β = 21.8, t = 21.22, p < .001, **Vowel/o/:** β = 9.6, t = 9.5, p < .001), in all places of articulation (see Table A6 for detailed results on interactions).

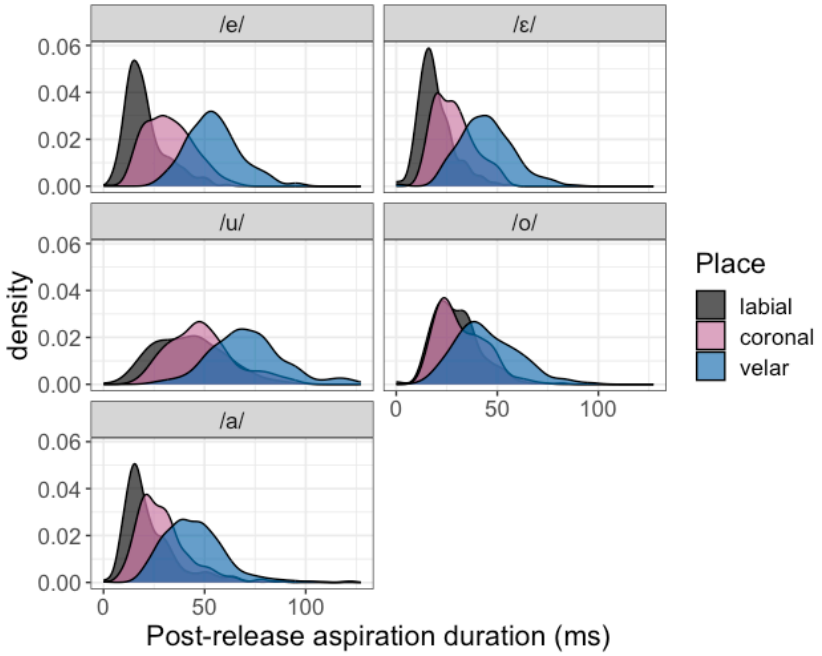


Figure 2.6 Density plot of post-release aspiration duration (ms) of all productions of voiceless stops. The Vowel /a/ is word-initial and the remaining vowels are word-medial.

Given the high prevalence of partially voiced *voiced* stops, it is pertinent to compare the post-release aspiration duration across voiced and voiceless stops. Figure 2.7 presents the distribution of post-release aspiration duration for voiced stops realized with incomplete closure voicing across all speakers. A linear regression, with the interaction of Place and Vowel and a by-subject random intercept, revealed the same overall patterns as for the voiceless stops, namely an increase in aspiration at more posterior places of articulation and for the vowels /u/ and /o/

(see Table A7). Similarly, a post-hoc analysis revealed no significant difference between labial and coronal stops preceding /o/ (β = 2.4, t = 1.8 p = .2), as in voiceless stops. A linear regression on the data subset containing voiceless stops and partially-voiced voiced stops with Voicing, Place, Vowel and the three-way interaction as fixed effects with a by-subject random intercept, reveals that post-release aspiration duration of voiceless stops is significantly longer than post-release aspiration duration of partially voiced *voiced* stops across all places of articulation and vowel contexts (all p's < .001), as seen in Table A8.

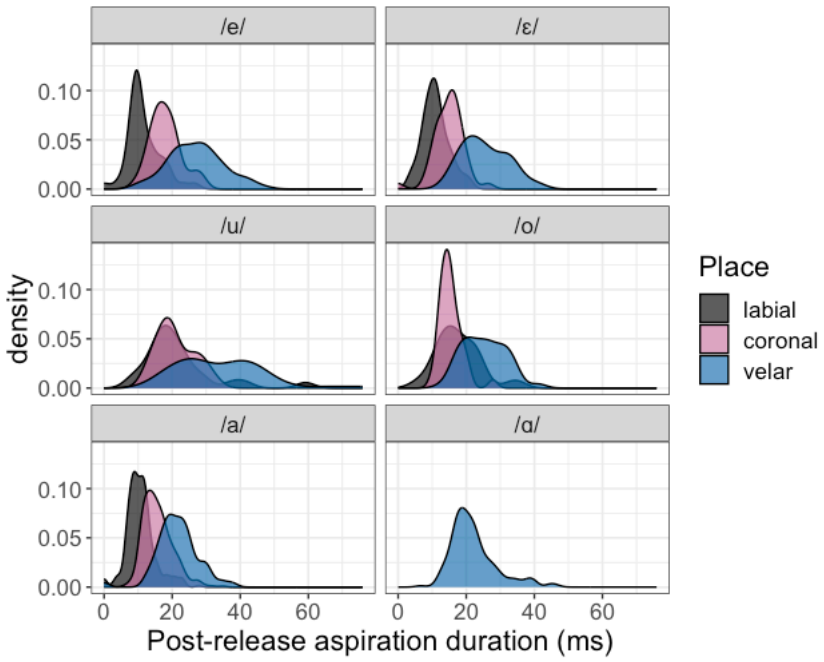


FIgure 2.7 Density plot of post-release aspiration duration (ms) of all productions of voiced stops realized with incomplete closure voicing. The vowels /a/ and /ɑ/ are word-initial and the remaining vowels are in word-medial position.

A considerable portion of fully voiced *voiced* tokens were also produced with a discernible burst and subsequent aspiration, as shown in Figure 2.3 (example B). Figure 2.8 shows the proportion of fully voiced stops produced with some aspiration. A linear regression with Place as a fixed effect and a by-subject random intercept revealed that the percentage of fully voiced stops produced with a burst and subsequent aspiration was significantly greater at more posterior places of articulation (**PlaceCoronal:** β = 30%, t = 9.8, p < .001; **PlaceVelar:** β = 44%, t =

14.7, $p < .001$). A greater prevalence of partial voicing accompanied by post-release aspiration is also observed in posterior voiced stops. Speakers are likely somewhat aware of this tendency and therefore expend less effort in ensuring voicing begins immediately following closure release, even when the entire closure is voiced. It is known that initiating and maintaining voicing at more posterior places of articulation is articulatorily more difficult because of the smaller supraglottal cavity. Instances of fully voiced stops produced with aspiration could also simply be tokens where the cessation of transglottal airflow coincided with the constriction release, resulting in a brief, unplanned period of voicelessness.

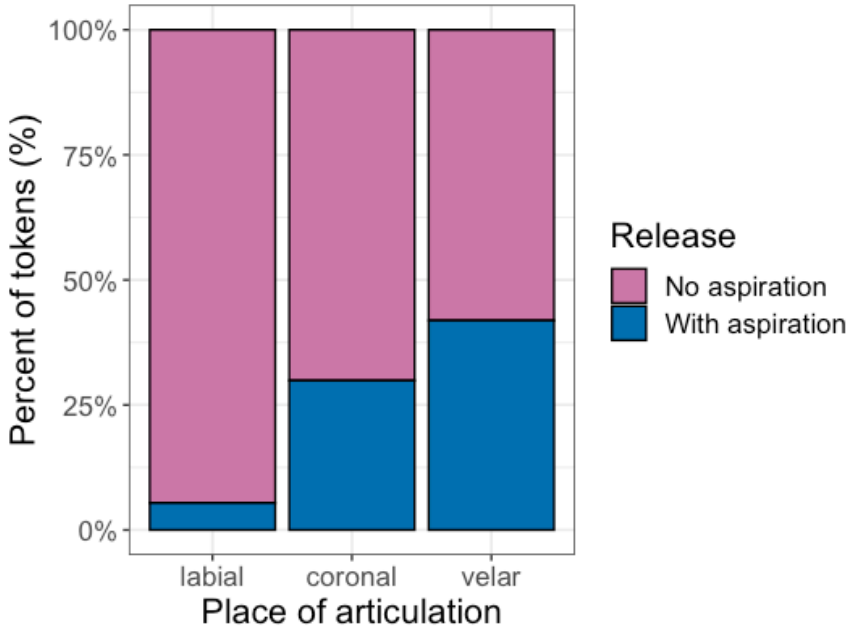


Figure 2.8 Proportion of voiced stops produced with full closure voicing either with or without post-release aspiration, corresponding to the voicing patterns (B) and (A), respectively, as defined in Figure 2.3.

2.2.1.3 Variability

As shown in the previous sections, a high level of variability in voicing patterns was observed across subjects. Figure 2.9 shows the distribution of voicing proportion for individual speakers. While many speakers produced nearly all voiced stops with full closure voicing (those on the far left of the figure), other participants produced all or nearly all voiced stops with wane

voicing (those furthest to the right). Carryover voicing in voiceless stops likewise exhibited a high degree of interindividual variability. Word position was not considered in the analysis of individual variability in order to better capture individual speaker trends. This was deemed appropriate given the lower token count for many speakers in word-initial stops (as a reminder, tokens produced with a discernible pause before the target word were excluded) as well as the importance of prioritizing differences across place of articulation.

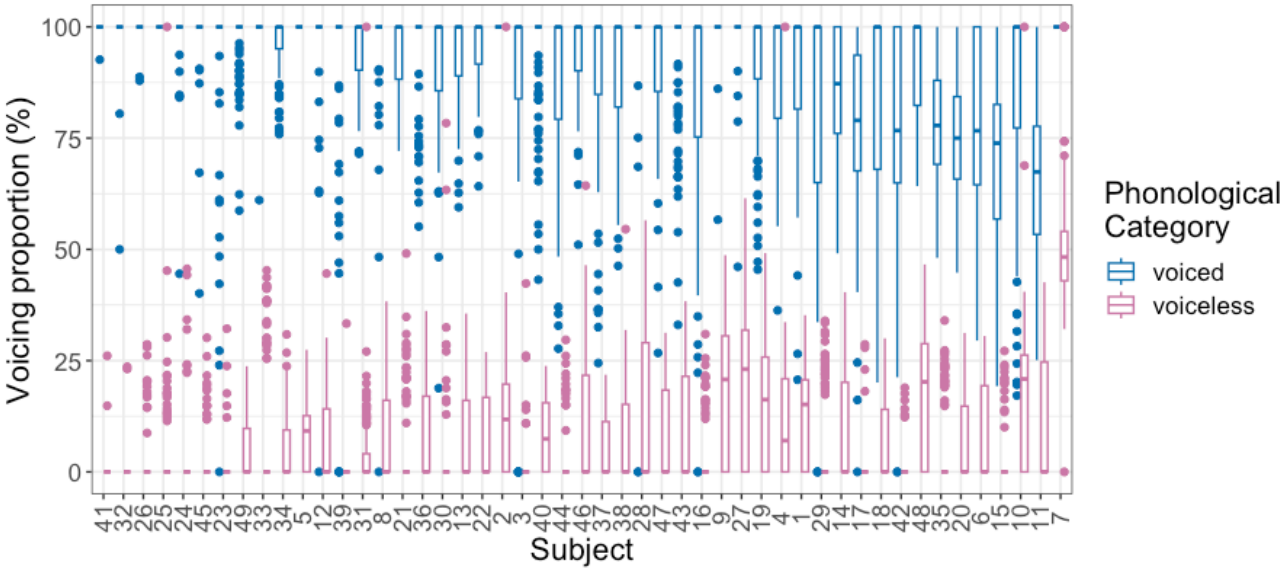


Figure 2.9 Boxplots of distributions of voicing proportion for voiced and voiceless stops, arranged left to right by the overall prevalence of partial voicing. Participants on the far left produced nearly all voiced stops with full closure voicing and all voiceless stops with no closure voicing. Participants on the far right either produced a large proportion of voiced stops with incomplete closure voicing or a large proportion of voiceless stops with carryover voicing. Note that the single dashes at 0% and 100% (for subjects 41 and 32, for example) are compressed boxes.

Although participants differ greatly from one another in how they produce voicing, the prevalence of partial voicing is highly consistent across different places of articulation within the speech of an individual. The mean voicing proportion of voiced and voiceless stops was calculated for each speaker and separately for each place of articulation. Pearson correlations across places of articulation, as seen in Figure 2.10, were highly significant. This implies that an individual who

consistently produced labial stops with full closure voicing, for example, also consistently produced velar stops with full closure voicing. Individual stop voicing behaviors therefore represent a consistent property of each speaker.

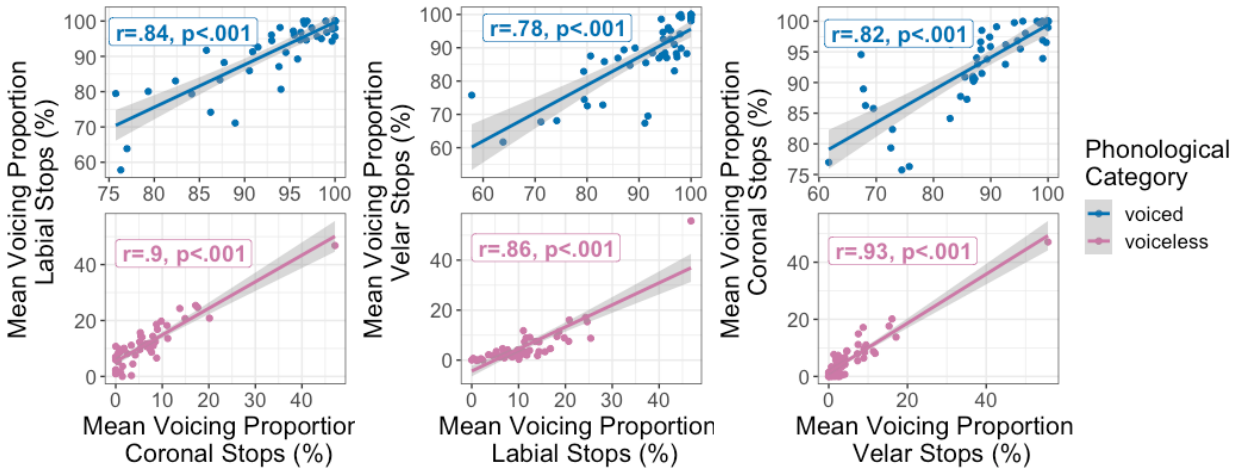


Figure 2.10 Pearson correlations between mean voicing proportion of voiced or voiceless stops across three places of articulation.

Similarly, participants varied considerably in their durations of post-release aspiration of voiceless stops, as evidenced in the distributions in Figure 2.6. However, all individuals produced highly consistent patterns across places of articulation, evidenced by the highly significant Pearson correlations in Figure 2.11.

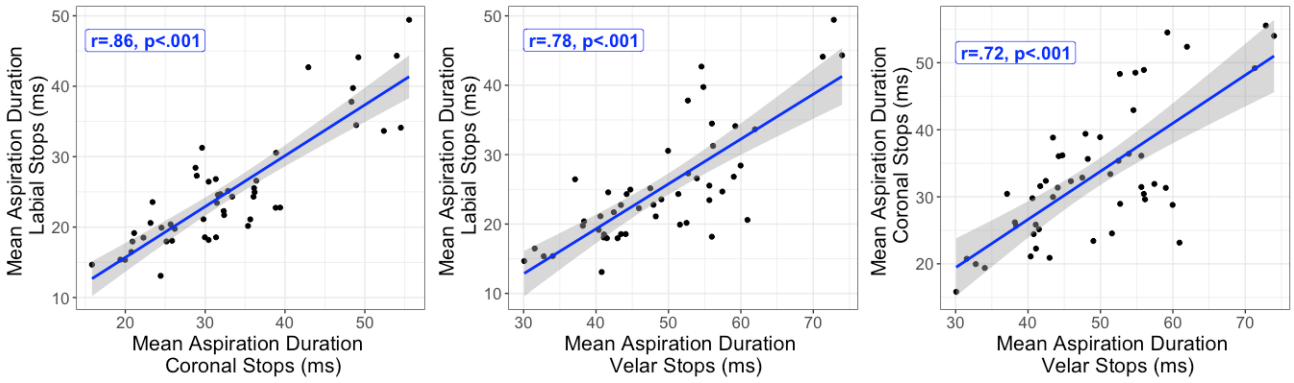


Figure 2.11 Pearson correlations between mean values of post-release aspiration duration of voiceless stops across three places of articulation.

To better quantify the amount of variability present in each participant's speech, Coefficient of Variation (CoV) values, one for voicing proportion and one for post-release aspiration duration, were calculated for each speaker at each place of articulation. A higher CoV indicates a wider distribution and therefore a higher degree of intraindividual variability. The Coefficient of Variation (CoV) is highly sensitive to outliers, making it ill-equipped to accurately represent variability in speakers who present skewed distributions of values, i.e., a mean close to zero with a few extreme outliers. For example, a speaker who only produced 2/35 labial voiceless stops with partial voicing would have a mean value of voicing proportion very close to zero (e.g., mean = 0.36 ms) but a much larger standard deviation (e.g., sd = 2.8 ms). This would lead to a very high CoV (e.g. CoV = 8), which is misleading seeing as this speaker produced voiceless stops with a high rate of consistency and therefore very little variability. Similarly, a speaker who only produced one or two voiced stops with post-release aspiration would have an artificially high CoV value for post-release aspiration duration, despite high consistency in voicing timing cues. For this reason, the CoV of voicing proportion was only calculated for voiced stops and the CoV of post-release aspiration duration was only calculated for voiceless stops, because within each of these two groupings, no speaker had a mean close to zero. Boxplots illustrating interindividual differences in intraindividual variability are shown in Figure 2.12, for voicing proportion (left) and post-release aspiration duration (right).

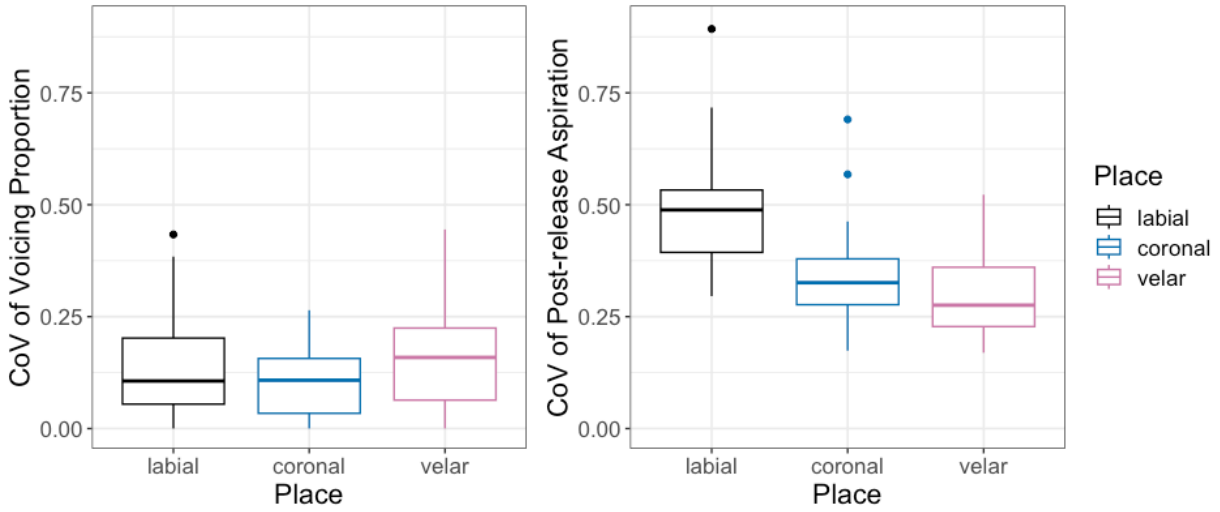


Figure 2.12 Distribution of Coefficients of Variation (CoV), representing intraindividual variability, of voicing proportion of voiced stops (left panel) and post-release aspiration duration of voiceless stops (right panel) across speakers.

Two linear regressions, one with CoV of post-release aspiration duration and the other with CoV of voicing proportion as dependent variables, with fixed effects of Place and a by-subject random intercept were carried out (see Table A9 and Table A10). The predictor Vowel was not included in these models because each participant produced only five tokens of each of the four word-medial vowels (/e, ɛ, o, u/) at each place of articulation. In addition, while participants produced a greater number of tokens of word-initial stops, preceding the vowel /a/, the distribution across place of articulation was unbalanced. The models confirmed that CoV of voicing proportion was lowest for coronal stops (β = -0.03, t = -2.6, p = .009) and highest for velar stops (β = 0.03, t = 2.3, p = .03). More variability in velar stops coincides with the observation that velar stops exhibited the highest percentage of both partial voicing and fully voiced stops realized with post-release aspiration. CoV of post-release aspiration duration decreased with more posterior places of articulation (**PlaceCoronal**: β = -0.14, t = -9.3, p < .001; **PlaceVelar**: β = -0.18, t = -12.09, p < .001). As a reminder, duration values of post-release aspiration were consistently longer at more posterior places of articulation, in the same context where variability is lowest. In other words, while all speakers consistently produced velar stops with longer durations of aspiration, speakers varied much more in their realization of labial stops.

Pearson correlation tests were carried out to determine if the rate of variability for an individual remains stable across places of articulation for a given acoustic property. All correlations of CoV values across pairs of places of articulation in both voicing proportion and post-release aspiration duration were significant, as seen in Figure 2.13. Correlations at the individual level of post-release aspiration duration (referred to as VOT in most studies) across stop categories have also been observed in other languages. Chodroff & Wilson (2017) reported strong correlations between means and standard deviations of VOT in English stops at different places of articulation. Hullebus et al. (2018) observed a strong correlation between standard deviations of VOT in German speakers' /t/ and /k/. Overall, these results offer strong evidence that variability across a single acoustic cue for segments belonging to the same larger phonological category, in this case voiced and voiceless stops, is highly consistent at the individual level and corresponds to a generalizable speaker trait.

The strong correlations of CoV across places of articulation provide empirical evidence that the Coefficient of Variation is an appropriate measure of intraindividual variability and that it is not artificially skewed by higher means. For example, the observation of similar degrees of variability of post-release aspiration duration across all three places of articulation suggests that, in general, an individual's standard deviation naturally increases as the mean value of that category increases. This proportional relationship is confirmed by significant correlations between the SD and the mean of the post-release aspiration duration for all three places of articulation (all r's > .4 and p's < .003).

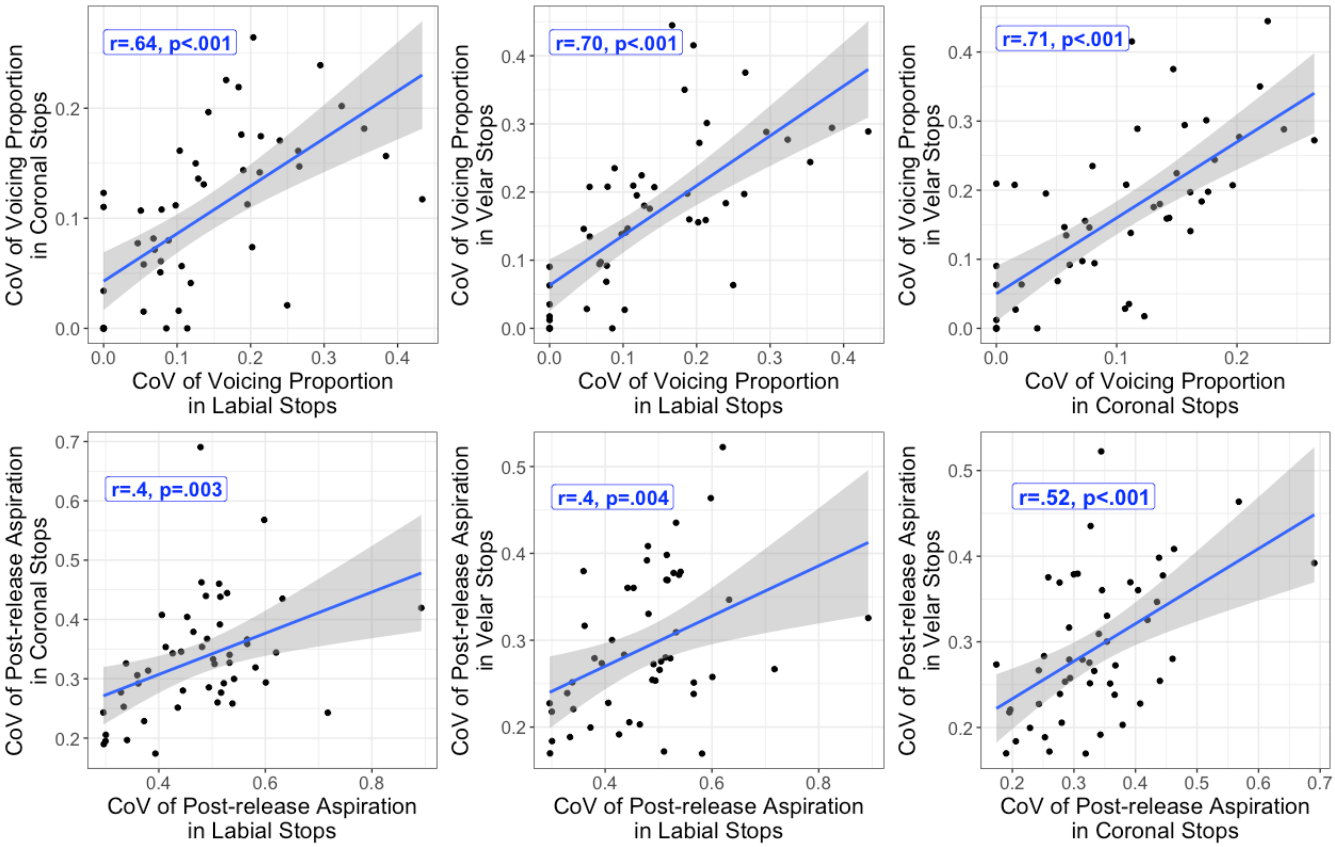


Figure 2.13 Pearson correlations of Coefficient of Variation (CoV) values of voicing proportion of voiced stops (top panel) and post-release aspiration duration of voiceless stops (bottom panel) across different places of articulation.

While rates of variability appear consistent across categories for the same acoustic measure, no correlation was observed between CoV of voicing proportion and CoV of post-release aspiration duration either for the same place of articulation or between different places of articulation ($p > .05$). Although both post-release aspiration duration and closure voicing are dependent on precise control of the vocal folds, the nature of their articulation is very different. Post-release aspiration refers to air passing through an open glottis that continues until the vocal folds are adducted to produce the voicing of the following segment (in the current study this was always a vowel). Closure voicing, however, requires the adduction of vocal folds and can only be initiated and maintained if the correct aerodynamic conditions are met (Ohala 1983). The presence of a correlation in intraindividual variability across places of articulation only for the same acoustic measure can be explained by shared acoustic and articulatory information.

Figure 2.10 and Figure 2.11 showed that mean values of partial voicing and post-release aspiration duration were also highly correlated across places of articulation. Therefore, the correlation of intraindividual variability is an indication that segments sharing similar ranges of a specific acoustic measure also have similar rates of variability. This connection is likely also due to the common articulatory configuration responsible for these acoustic values. While the articulatory configuration necessary to produced voiced and voiceless stops differs greatly, within each of these two phonological categories, the configuration remains largely the same across different places of articulation. Distinct strategies can be employed in the production of both voiced and voiceless stops to facilitate or prevent voicing, respectively, such as larynx lowering/raising, tongue root advancement/retraction, tongue body lowering/raising, etc., but the principal articulator are always the vocal folds (Perkell 1969, Westbury 1983, Ahn 2018a, Ahn 2018b, Zhang & Goldstein 2023, Schweizer et al. 2025). Therefore, the high correlations in variability can be seen as stemming from a common distinct glottal configuration employed in the realization of a category of segments (here two distinct configurations for voiced and voiceless stops), regardless of the area of constriction in the oral cavity. Given the highly consistent rates of intraindividual variability across segments, the idea that higher variability in acoustic values is an indication of a lower degree of precision in articulation is not necessarily well-founded. Such variability may not be representative of failed articulatory maneuvers or a lack of articulatory effort, but rather the result of a stable speech production system composed of a wider range of acceptable articulatory configurations.

2.2.2 Vowels

The data for the vowel analysis was constructed as follows. First, all instances of vowels following the palatal nasal stop /ɲ/ were removed due to the difficulty of distinguishing between and segmenting the two main variants observed: /ɲ/ and /nj/. Additionally, the final vowels in the words *bazou* /bazu/ and *gager* /gaʒe/ were removed, given that only one word with each of the word-medial consonants /z/ and /ʒ/ was recorded. F1 and F2 values were extracted at every millisecond of the vowel using PraatSauce (Kirby 2018). To remove formant-tracking errors, all

formant values were first transformed into z-scores by speaker and abrupt jumps in formant tracking were removed by excluding values that differed by more than 0.05 in z-scores. Then, the steady-state portion (40%-60%) of the vowel was isolated in order to avoid influence from onset or offset transitions. After this, for each vowel and speaker, values that were two or more standard deviations away from the mean were excluded. Overall, this removed approximately 4% of formant values (F1: n=324/8879; F2: n=314/8879). Finally, to obtain a single formant measure for each speaker, the formant values in the steady-state portion of the vowel were averaged together. These mean values were used in all subsequent analyses. To allow comparison across speakers, formant values were normalized by means of newly calculated by-subject z-scores. These z-scores were subsequently converted back onto a recognizable scale by multiplying them by the aggregated standard deviation and adding them to the aggregated mean, following Brunelle et al. (2020). These normalized values were only used when analyzing the formant values in Hertz and not when analyzing the individual variability.

2.2.2.1 Vowel formants (F1 and F2)

Unsurprisingly, vowels were realized in a large range of F1-F2 space with considerable interindividual variability. Overall, the front vowels /e/ and /ɛ/ were produced with less overlap than the back vowels /u/ and /o/, as shown in Figure 2.14 and Figure 2.15. Given the large degree of variability observed, specifically in back vowels, as well as the visible overlap in F1 and F2 values for some speakers, as shown in Figure 2.14, Pillai scores were calculated to determine the degree of distinction between the two vowel pairs occupying similar acoustic space, namely /e-ɛ/ and /u-o/. Pillai scores are calculated by means of multivariate analyses of variance (MANOVA) which consider two dependent variables, in this case F1 and F2, to determine if there exists a significant difference between two groups, in this case two vowel categories (Hay et al. 2006a). The resultant Pillai-Bartlett trace, most often referred to as Pillai score, is a very common metric used specifically in sociolinguistics to investigate vowel mergers and contrasts (Sloos 2013; Kendall & Fridland 2017, among many others). A Pillai score ranges from 0 to 1 where a lower score is taken as evidence of a vowel merger and a higher score indicates completely distinct

categories in production. The literature has yet to agree upon a standard cut-off value to decide if two vowel categories are merged. Stanley & Sneller (2023) suggest determining the merger value as a function of sample size, where larger samples would require a much lower Pillai score to consider the vowels merged. Following Stanley & Sneller (2023), the sample size in the current study would require a Pillai score of less than 0.1 to indicate a true vowel merger.

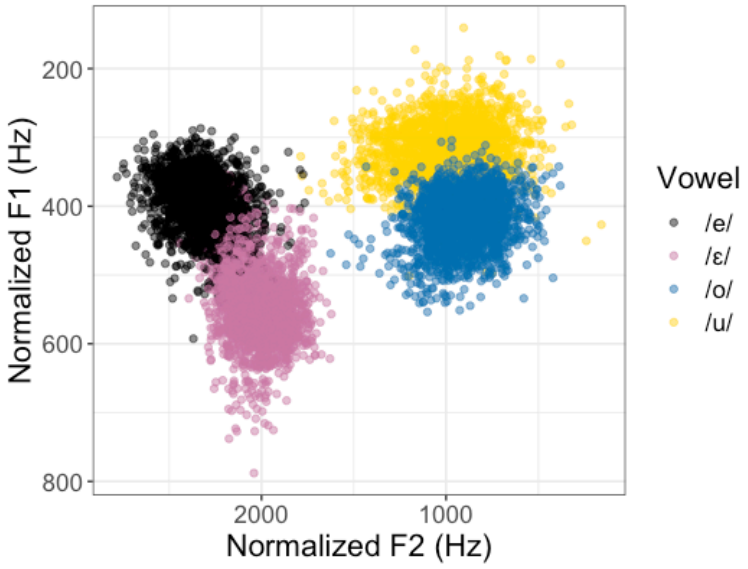


Figure 2.14 Scatterplot of normalized F1 and F2 values, representing the overall vowel space for the four vowels recorded in Production B.

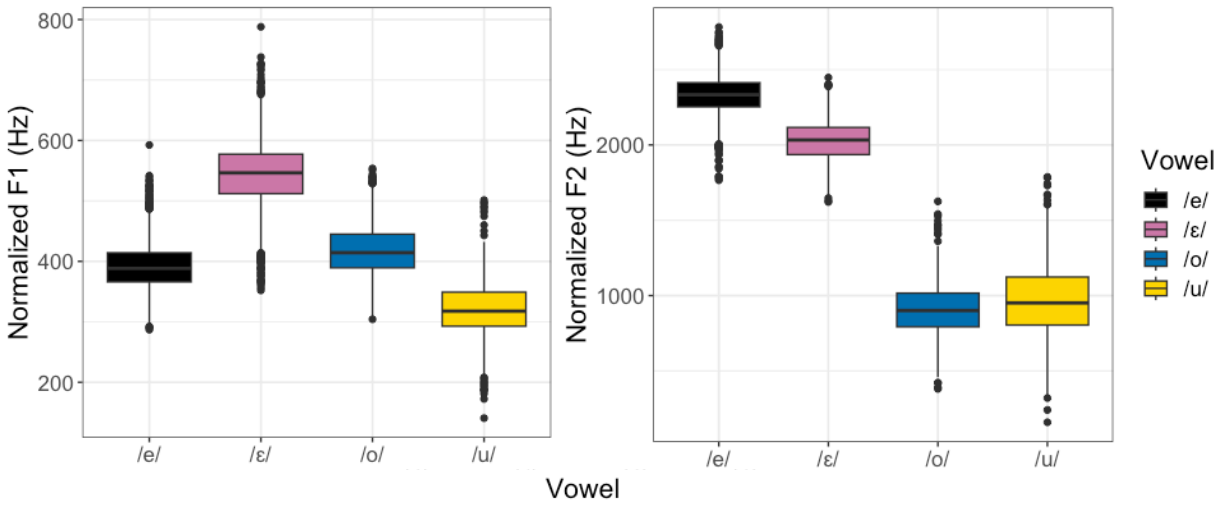


Figure 2.15 Boxplots of normalized F1 and F2 values in Production B.

Figure 2.16 presents the Pillai scores for the two vowel contrasts for each speaker, arranged on the x-axis from lowest to highest Pillai score for the /u-o/ contrast. It is immediately clear that all vowels are distinct for all speakers, although, as expected, considerable interindividual variability is present. The /e-ɛ/ contrast is consistently produced as more distinct than the /u-o/ contrast, with two or three exceptions. The fact that a few speakers obtained a Pillai score of less than 0.5 for the /u-o/ contrast is indicative of some overlap, similar to what has been observed for the prevelar /æ-ɛ/ contrast in some dialects of North American English (Freeman 2023).

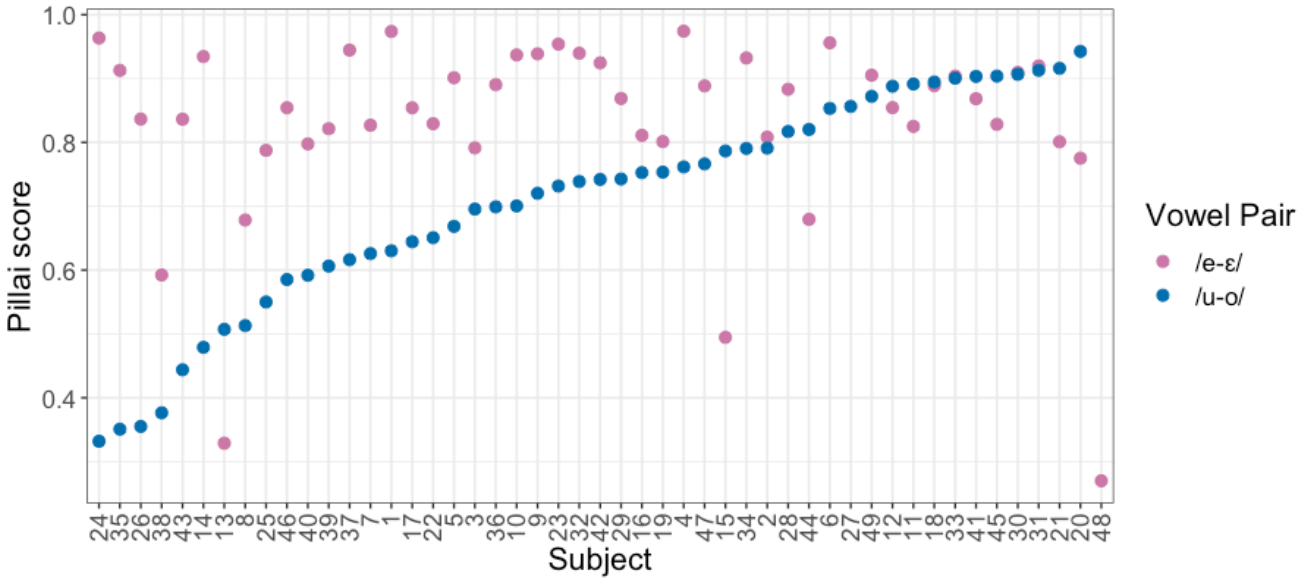


Figure 2.16 Individual speaker Pillai scores for the two vowel contrasts studied. A lower Pillai score indicates more overlap in vowel categories. Only statistically significant Pillai scores are shown.

However, the distinction between vowel pairs is likely underestimated given that additional acoustic properties that are known to be contrastive, such as vowel length or rounding, were not considered here. Moreover, the degree of distinction between phonemes has also been shown to be partially dependent on the functional load of the phonological contrast in question. In other words, if there are more minimal pairs in the lexicon that differ exclusively along a specific vowel dimension, the contrast between the two vowels is likely to remain more distinct (Wedel et al. 2013; Stevenson & Zamuner 2017). In Laurentian French, the /e-ɛ/ contrast is very

common, largely due to its role in distinguishing the future and conditional endings. Given this, the stronger distinction observed between this vowel pair could be due to the greater functional load within the lexicon. While minimal pairs also exist for the /u-o/ contrast, they are arguably less salient, which could reduce the need for speakers to maintain such distinct categories. However, given that both contrasts consist of core phonemes of Laurentian French, it is likely that any effect functional load may have is minimal.

2.2.2.2 Variability

As for stop voicing, variability in F1 and F2 was quantified by means of the Coefficient of Variation (CoV) calculated separately for each speaker and each vowel, shown in Figure 2.17. What is immediately apparent is the lower values of CoV in comparison to those seen for stops. As a reminder, the CoV of voicing proportion ranged between 0 and 0.4, while the CoV of post-release aspiration duration ranged between 0.2 and 0.8. To quantify differences in CoV across vowels, two separate mixed-effect linear regressions were run, one for CoV of F1 (Table A11) and one for CoV of F2 (Table A12), both with Vowel as a fixed effect and a random by-subject intercept. Post-hoc analyses revealed that CoV of F1 was significantly greater for /u/, compared to /e/ (β = 0.02, t = 3.28, p = .007) and /o/ (β = 0.02, t = 3.8, p = .001). CoV of F2 was significantly higher for both /u/ (β = 0.16, t = 27.9, p < .001) and /o/ (β = 0.1, t =17.6, p < .001) in comparison to /e/. Post-hoc analyses confirmed CoV of F2 of both back vowels was significantly greater than CoV of F2 of /ɛ/ as well (**Vowel/u/:** β = 0.1, t = 17.63, p < .001; **Vowel/o/:** β = 0.1, t = 17.1, p < .001). CoV of F2 of /u/ was also significantly greater than that of /o/ (β = 0.06, t = 10.3, p < .001). The increased variability in back vowels is possibly due to the role of formant transitions. Namely, F2 onset in post-coronal back vowels is much higher than that of front vowels, and the most represented place of articulation in the dataset was coronal (oral stops, fricatives and nasal stop). Although the formant measures were taken from the middle portion of the vowel, coarticulatory effects may extend past vowel onset and affect the formant structure even during the purported steady state.

An additional observation is that more variability was observed in vowel height (F1) for front vowels and vowel backness (F2) for back vowels. This was confirmed by an additional mixed-effect linear regression with Vowel and CoV Type as fixed effects and a by-subject random intercept (Table A13). Post-hoc analyses on this regression confirmed that CoV of F1 was greater than CoV of F2 for front vowels, whereas the opposite was true in back vowels (**Vowel/e/**: β = 0.01, z = 2.8 p = .005; **Vowel/ɛ/**: β = 0.02, z = 3.9 p < .001; **Vowel/o/**: β = -0.9 z = -16.8 p < .001; **Vowel/u/**: β = -0.12 z = -24.1 p < .001). The observation of a higher degree of variability in the F2 of back vowels, which naturally have a higher mean than the F1 of back vowels, offers support that the Coefficient of Variation is an accurate measure, given that this observed tendency is opposite to what would be expected if greater means inherently resulted in lower CoVs.

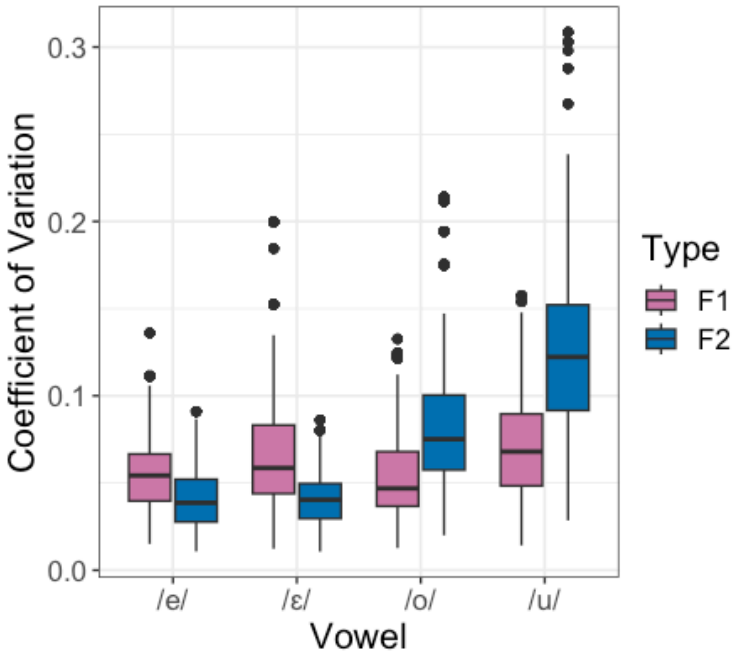


Figure 2.17 Distribution of Coefficients of Variation (CoV), representing intraindividual variability, of F1 and F2 values across all speakers.

In order to assess the generalizability or specificity of individual variability, two sets of Pearson correlations were carried out. First, the relationship between CoV values of F1 and F2 within the same vowel was analyzed. All four vowels showed highly significant yet moderate

correlations between CoV of F1 and CoV of F2, as illustrated in Figure 2.18. This is not particularly surprising, given that F1 and F2 values themselves are often correlated, and the division of vowel production into distinct values for horizonal and vertical tongue position is an oversimplification. Indeed, although the categorization of vowels according to binary divisions of tongue height and backness has proven fruitful in understanding certain phonological processes, such as vowel harmony or vowel mergers, each vowel production represents a unique tongue shape.

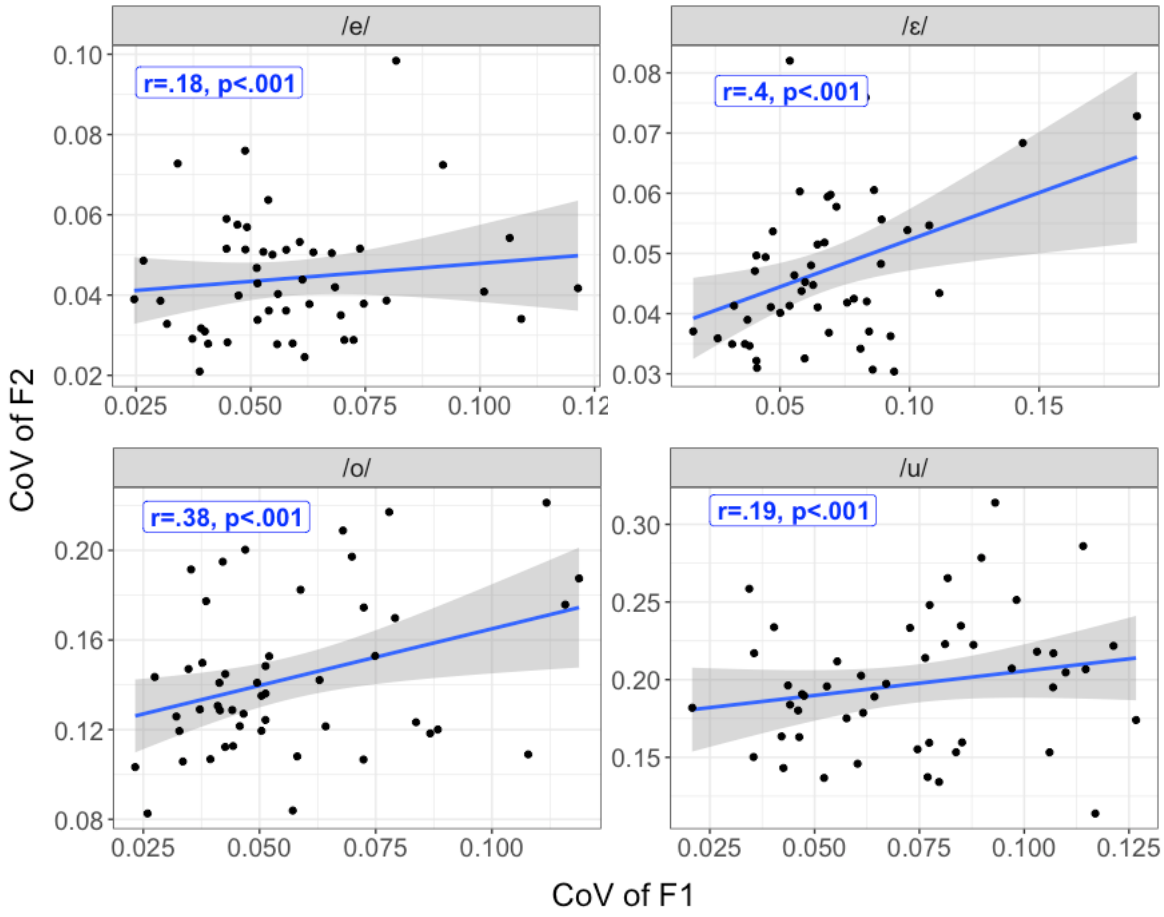


Figure 2.18 Significant Pearson correlations between the Coefficient of Variation of F1 and the Coefficient of Variation of F2 within the same vowel.

To evaluate if variability is specific to each vowel or indicative of an overall speech trait, CoV values of F1 and CoV values of F2 were compared across the six possible vowel pairings. Figure 2.19 shows the only significant correlation for CoV of F1, between /e/ and /o/, and a non-significant but trending correlation for CoV of F1 between /ɛ/ and /o/. All other correlations were

non-significant (p > .2). Given that /e/ and /o/ are produced with highly similar F1 values, the correlation in the amount of variability observed could be representative of variability in control of a specific vertical lingual dimension. This interpretation aligns with the observation that the correlation between CoV values in voiced and voiceless stops across distinct places of articulation is likely due to highly similar articulatory configurations of the vocal folds. The lack of significance of the CoV of F1 correlation between /ɛ-o/ would then be explained by a reduced similarity in vowel heights. Whalen et al. (2018) found a similar result in their analysis of variability in English vowels, namely that the only vowel pairs with correlated acoustic variability were /ɛ-ʌ/ and /ʌ-ɔ/, all of which are produced with similar F1 values.

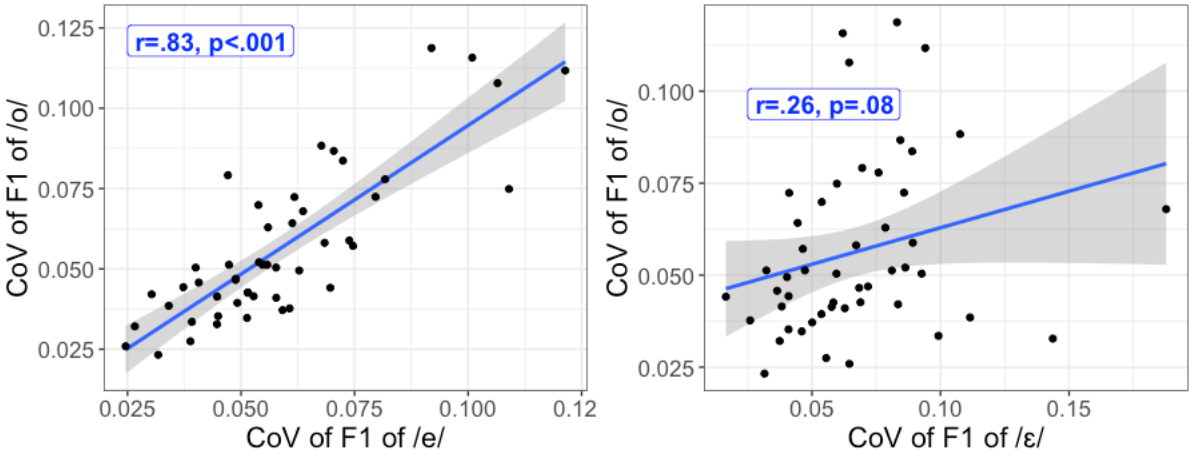


Figure 2.19 Significant Pearson correlation between the Coefficients of Variation of F1 of the vowel pairs /e-o/ and /ɛ-o/.

Figure 2.20 shows all correlations between the CoVs of F2 across the six vowel pairings. Four comparisons were significant (/o-u/, /o-e/, /o-ɛ/ and /e-ɛ/), although with varying effect sizes. No significant correlation was observed between /u-ɛ/ and /u-e/. Both vowels in the pairs /o-u/ and /ɛ-e/ are produced with highly similar F2 values, again suggesting that variability is likely correlated across segments only when the acoustic values themselves are similar, as for F1 across the vowel pair /e-o/. The lack of correlation between /u/ and the front vowels /e/ and /ɛ/ is then explained by the large difference in F2 values, implying reduced articulatory similarity. However, there are certain inconsistencies. First, a significant correlation was observed between the CoV of F2 of the vowels in the pairs /o-ɛ/ and /o-e/, despite these vowels contrasting primarily

along the F2 dimension. Given the highly significant correlations between CoV of F1 and F2 values within each individual vowel, shown in Figure 2.18, it is possible that the significant correlation between F1 values of the vowel pair /o-e/ and the trending correlation of /o-ɛ/ are also reflected in the CoV of F2 values. In addition, the tongue being a muscular hydrostat, any change in one dimension of tongue position, i.e. vertical or horizontal, naturally incurs a change in the other dimension (Kier & Smith 1985). It could also simply be the case that certain idiosyncratic relationships exist across vowels and that correlations in variability may not depend exclusively on the degree of similarity in acoustic and/or articulatory dimensions. Additionally, the potential confounding effects of the Coefficient of Variation, as discussed in Section 2.1.5, may play some role in the observations shown here. Specifically, the fact that mean F2 values of front vowels are naturally higher than those of back vowels, could indicate that the degree of variability of certain vowels was inaccurately represented. Nevertheless, the overall results clearly suggest that, at some level, similarity in articulation likely plays a role in the degree of intraindividual variability in vowels.

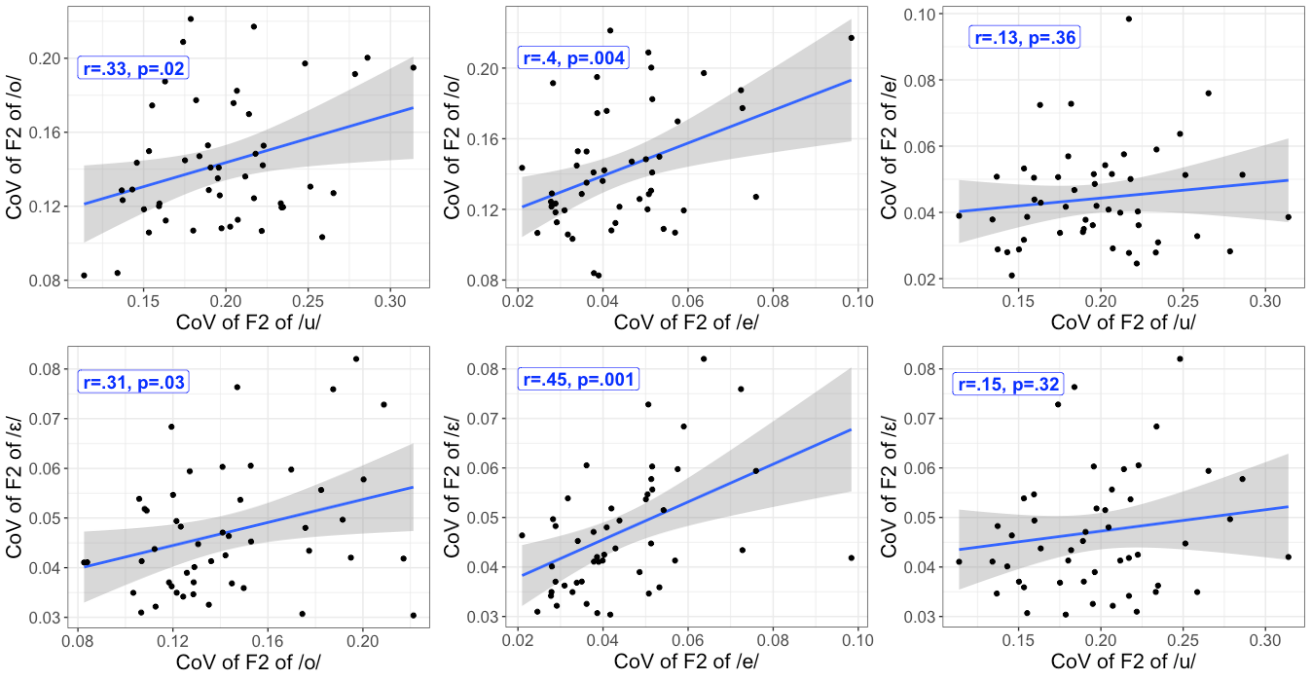


Figure 2.20 Pearson correlations between the Coefficients of Variation of F2 of distinct vowel pairs.

2.2.3 Fricatives

As described in Section 2.1.4, the first four spectral moments of the fricatives /s/ and /ʃ/ were extracted. These four spectral moments correspond to the center of gravity (COG), standard deviation, skewness and kurtosis. To avoid the effect of on- and off-transitions, average values were taken from the middle 80% of frication by means of a Praat script (DiCanio 2013). To remove measurement errors, a similar approach was used as for the vowels. Means and standard deviations were calculated for each fricative and speaker and values that were two or more standard deviations away from the mean were excluded. Overall, this removed 4% of COG values (n=69/1904), 5% of standard deviation values (n=91/1904), 5% of skewness values (n=90/1904) and 8% of kurtosis values (n=164/1904). These spectral measures were normalized following the same procedure as for the vowels, namely by means of by-subject z-scores. These z-scores were then converted back onto the original scale by multiplying by the aggregated standard deviation and adding the aggregated mean. These normalized values are only used in the figures depicting aggregated values of the spectral moments and in the statistical models, but not in the analysis of intraindividual variability. Mixed-effect models were built to test the relevance of each spectral cue in distinguishing the two fricatives across speakers. All models included FRICATIVE (**/s/**, /ʃ/) as the fixed effect with a by-subject random intercept.

2.2.3.1 Spectral moments

As evidenced in Figure 2.21, despite a considerable degree of interindividual variability, the fricatives were consistently distinguished across all four spectral moments, that is, COG (β = -2598, t = -90.7, $p < .001$), standard deviation (β = -500, t = -36.1, $p < .001$), skewness (β = 1.4, t = 43.7, $p < .001$) and kurtosis (β = -0.8, t = 7.2, $p < .001$). Nevertheless, there remains a high degree of variability especially in kurtosis, for which normalized values present a near complete overlap.

To assess the relative importance of each acoustic measure on the individual level, Cohen's ds were calculated. Cohen's ds can be used to determine the magnitude of differences between two distributions, corresponding here to the spectral values between /s/ and /ʃ/ (Cohen

1998). The coefficient is calculated for each individual by dividing the mean difference between the two categories by the combined standard deviation of the two categories. Cohen's ds are presented in Figure 2.22, where a positive value indicates that a higher value for that spectral measure is associated with production of /s/. For nearly all speakers, COG had a much greater effect size than the remaining three spectral moments, although the relevant weight of each cue varies considerably across speakers.

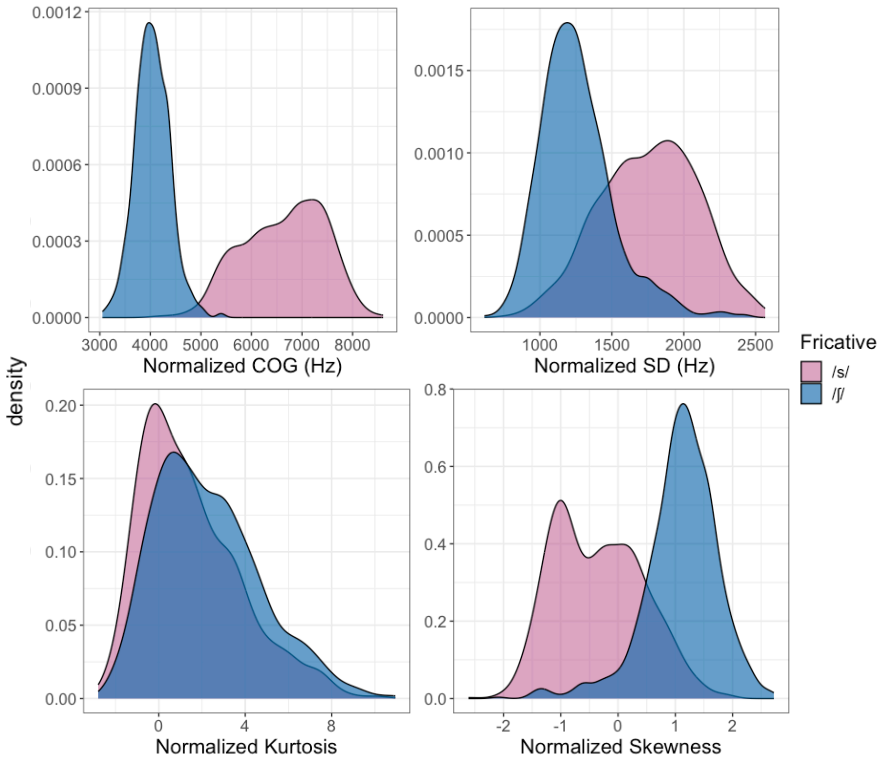


Figure 2.21 Density plots of normalized values for the first four spectral moments of the fricatives /s/ and /ʃ/.

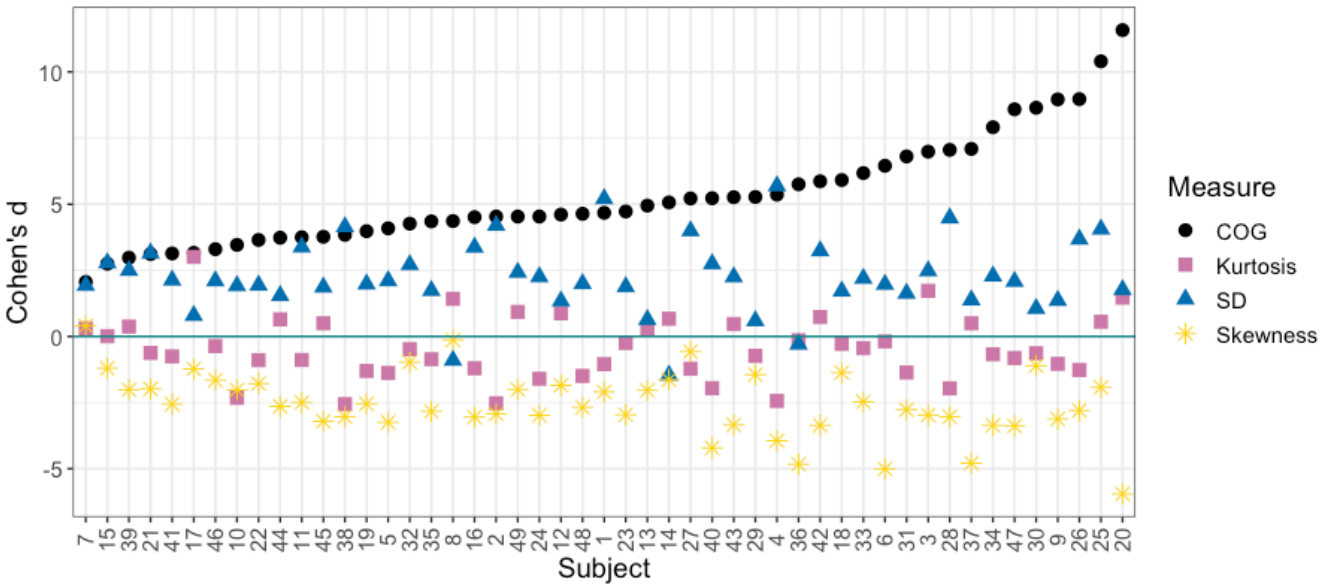


Figure 2.22 Speaker-level Cohen's d values for the first four spectral moments of /s/ and /ʃ/, ordered from lowest (leftmost) to highest (rightmost) Cohen's d for COG.

2.2.3.2 Variability

Given the overwhelming importance of COG for all speakers, evidenced by the larger absolute Cohen's d observed in Figure 2.22, variability in fricatives was analyzed solely for COG. Figure 2.23 shows the distribution of Coefficient of Variation (CoV) values for the two fricatives. A linear regression with Fricative as the fixed effect and a random by-subject intercept confirmed that more variability was observed in /s/ than in /ʃ/ ($\beta = 0.04$, $t = 6.5$, $p < .001$), aligning with Harper (2021) who also observed greater variability in /s/ than /ʃ/ in English speakers.

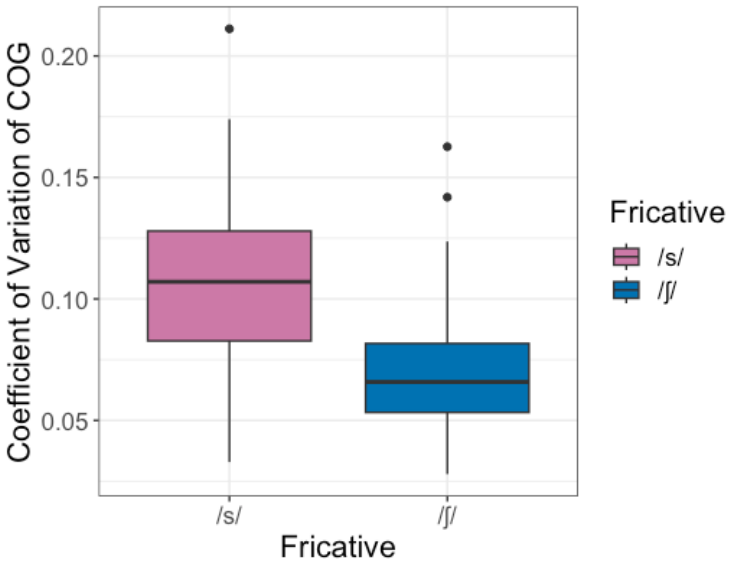


Figure 2.23 Distribution of Coefficients of Variation (CoV), representing intraindividual variability, of Center of Gravity (COG) of fricatives across all speakers.

A Pearson correlation revealed no significant relation between CoV of COG across fricatives, as shown in Figure 2.24, also coinciding with Harper (2021)'s observation of a lack of correlation between CoV of COG values of the /s-ʃ/ contrast in English. The observation of a greater degree of variability for /s/, which has a mean value much higher than that of /ʃ/, rules out the potential confound of an artificial decrease of CoV at higher means. This serves as additional evidence that the CoV is an accurate measure of variability and that a higher mean does not inherently lead to a lower CoV.

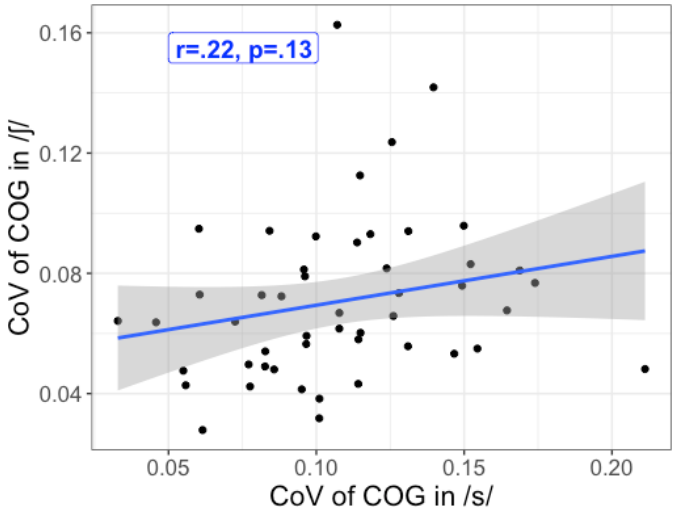


Figure 2.24 Pearson correlation between the Coefficient of Variation of COG in /s/ and /ʃ/.

Unlike stop voicing across different places of articulation and vowel pairs with similar tongue positions, /s/ and /ʃ/, while both coronal consonants, are produced with distinct tongue shapes. Although considerable interspeaker variability has been reported, /s/ is generally produced with the tongue tip while /ʃ/ is produced with the tongue blade in French (Toda 2009) and English (Narayanan et al. 1995). Additionally, articulatory studies have revealed that the tongue body is flatter in /s/ than in /ʃ/ in French (Toda 2009) and English (Narayanan et al. 1995). These articulatory differences are responsible for the significant differences in COG values across the two fricatives, which is in turn responsible for upholding the phonological contrast. If variability is indeed correlated only for segments with similar acoustic and/or articulatory values, it logically follows that individuals produce each fricative with a different rate of variability.

2.2.4 Variability across segments

To assess the comparability of intraindividual variability across distinct phonological segments and acoustic measures, a single mean CoV value was calculated for each acoustic measure and speaker, averaging across different categories, such as Vowel or Place. This resulted in five distinct CoV measures: voicing proportion, post-release aspiration duration, F1, F2 and COG. For the most part, the degree of variability differed more across acoustic categories than

across individual speakers, as visualized in Figure 2.25. As seen in previous sections, the highest degree of variability was observed in the realization of voiced and voiceless stops. For nearly all speakers, post-release aspiration duration was the most variable property, followed by proportion of voicing. The remaining measures, F1, F2 and COG, were realized with highly similar degrees of variability across speakers, although interindividual differences are still present.

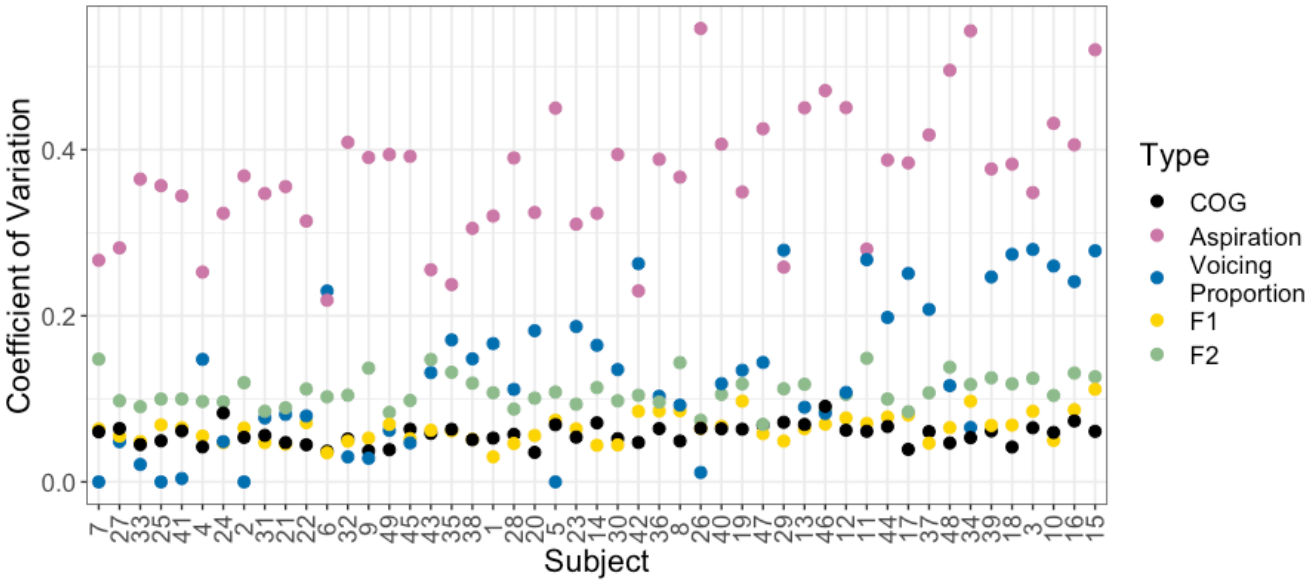


Figure 2.25 Individual mean Coefficient of Variation (CoV) values for each relevant acoustic measure, arranged by the total sum of CoV measures (lowest to highest from left to right).

In order to evaluate if the degree of variability represents a generalizable characteristic of the speaker or is dependent on the phonological category and acoustic property measured, Pearson correlations were carried out for all combinations of the five CoV measures. As shown in Table 2.5, the only significant correlation was between CoV of F1 and CoV of F2. Given that vowel formants are largely determined by tongue position, and any change in one spatial dimension of the tongue, i.e. horizontal or vertical, will necessarily cause an opposite change in the other dimension (Kier & Smith 1985), it is unsurprising that rates of variability are somewhat correlated. Nevertheless, the correlation is only moderate and likely still indicates some interdependence of tongue height and backness. The observation that no other correlation was

significant emphasizes that variability for a given acoustic measure is largely an independent property of each individual speaker.

Table 2.5 Pearson correlations between the Coefficient of Variation of five acoustic measures. A single asterisk indicates a significance of p = .01. *AspDur* indicates post-release aspiration duration and *VoiProp* indicates voicing proportion.

	CoV AspDur	**CoV VoiProp**	**CoV F1**	**CoV F2**	**CoV COG**
CoV AspDur		r = -.21	r = .24	r = -.19	r = -.20
CoV VoiProp	r = -.21		r = .16	r = .19	r = -.02
CoV F1	r = .24	r = .16		**r = .28***	r = .12
CoV F2	r = -.19	r = .19	**r = .28***		r = -.02
CoV COG	r = -.20	r = -.02	r = .12	r = -.02	

2.3 Interim summary and discussion

Intraindividual variability differed considerably both across speakers and acoustic measures. Overall, a speaker who produces one category with a high degree of variability does not necessarily possess more variable speech overall. Although certain idiosyncrasies were observed, the general tendency is that an individual speaker's rate of variability is consistent across similar acoustic values likely resulting from similar articulatory configurations. Variability in voicing proportion across different places of articulation was highly correlated. Similarly, variability of post-release aspiration duration was highly correlated across different places of articulation. However, no correlation was observed between variability of these two acoustic measures. Similarly, a correlation was observed between the F2 variability in both the back vowel pair /u-o/ and the front vowel pair /e-ɛ/, which share similar acoustic ranges. However, variability in F2 was also found to be correlated across /o-e/ and /o-ɛ/, despite these two pairs differing considerably in F2 values. It was discussed that this is possibly a 'side effect' of the similarity in F1 values, given that the horizontal and vertical dimensions of the tongue are not independent

of one another. Regarding F1, /o-e/ was the only vowel pair for which a significant correlation in variability was found. Average F1 values for these vowels are nearly identical, representing highly similar vertical tongue placement. For fricatives, no correlation was observed between CoV of COG of /s/ and CoV of COG of /ʃ/. Given the consistently distinct COG measures across the two fricatives, the lack of variability corroborates the idea that variability is consistent only within similar ranges of acoustic values. Additionally, as mentioned, these fricatives are produced with distinct tongue configurations (Toda 2009), implying variability is a shared characteristic across similar articulatory configurations as well as their produced acoustic ranges.

These results align with those of Harper (2021), who observed a correlation between the variability of tongue tip position of /s/ and /t/, which have very similar articulatory configurations. However, the variability of other articulatory markers, such as tongue back or lip movement, was uncorrelated, both within a single segment (tongue tip of /s/ and tongue back of /s/) and across segments (tongue back of /s/ and tongue back of /t/). Likewise, Harper (2021) found no correlation in formant variability between English /l/ and /ɹ/, likely due to the large difference in tongue shape required for producing these sounds, despite both often being categorized as coronal.

If variability is indeed specific to a precise articulatory configuration, rather than a generalizable characteristic across phonological segments (such as all vowels or all fricatives) or across similar acoustic outputs (such as formants or COG), the following predictions can be made that can be tested with the data in this dissertation. Although the following analyses were not initially intended, the carefully designed word list and controlled recording and analysis procedure allows for additional measures to be extracted from the acoustic data collected in Production B.

1. **Correlation** in the variability of f0 across different vowels following voiced and voiceless stops, given that f0 is primarily determined by a similar configuration of the adducted vocal folds. Voiced and voiceless stops were analyzed separately given that f0 in vowels following voiceless stops has been shown to be higher than that following

voiced stops, which could indicate a distinct articulatory mechanism (Jacques & Charbonneau 1989).

2. **Correlation** in the F1 variability of /y/ (found in the frame sentence preceding each target word) and /u/, given that both vowels have highly similar F1 values and are produced with similar tongue heights.
3. **Possible correlation** in the F2 variability of /a/ and /ɛ/, given that these vowels have similar F2 values and are produced at a similar horizontal tongue locus.
4. **No correlation** between the F2 variability of /y/ and that of any other vowel, given that /y/ is produced more fronted than the other vowels in this study.
5. **No correlation** between the F1 variability of /a/ and that of any other vowel, given that /a/ is the only low vowel in the study.

F0 values and formant values for /y/ and /a/ were extracted along with all other acoustic measures by means of PraatSauce (Kirby 2018) and tracking errors were removed following the same procedure as described in Section 2.1.4. As a reminder, the vowel /y/ was present in every token within the carrier phrase *T'entends-tu* _____ *pour vrai?* /tɑ̃tɑ̃-t**y** ____ puʀ vʀɛ/ 'Do you hear ____ for real?'. All instances of devoiced /y/ were excluded. The vowel /a/ was the vowel of the first syllable of every target word, except for the two words with the back vowel /ɑ/, which were not included in this analysis. Coefficient of Variation (CoV) values were calculated for these additional segments to determine the degree of variability.

The results from the Pearson correlations show that all predictions are borne out as presented in the following figures. Table 2.6 shows very strong correlations in variability of f0 in all environments, indicating that variability in f0 affects an individual's speech across the board. Figure 2.26 shows a strong correlation between CoV of F1 for the vowel pair /y-u/ and a weaker correlation between CoV of F2 for the vowel pair /a-ɛ/, corresponding to predictions 2 and 3 above. Figure 2.27 presents the normalized F1 and F2 values of these two additional vowels alongside the four vowels of interest in the study to illustrate the similarities in acoustic values across the vowel pairs.

Table 2.6 Pearson correlations of the Coefficients of Variation of f0 across four vowels following voiced and voiceless stops. All correlations are highly significant, indicated by the asterisk ($p < .000001$). *Vcd* indicates the vowel was preceded by a voiced stop and *vls* indicates the vowel was preceded by a voiceless stop.

	/e/_vcd	**/u/_vls**	**/u/_vcd**	**/e/_vls**	**/ɛ/_vcd**	**/o/_vls**	**/o/_vcd**	**/ɛ/_vls**
/e/_vcd		r = .66*	r = .83*	r = .89*	r = .84*	r = .89*	r = .85*	r = .86*
/u/_vls	r = .66*		r = .81*	r = .60*	r = .54*	r = .71*	r = .70*	r = .62*
/u/_vcd	r = .83*	r = .81*		r = .77*	r = .76*	r = .86*	r = .85*	r = .78*
/e/_vls	r = .89*	r = .60*	r = .77*		r = .85*	r = .86*	r = .82*	r = .84*
/ɛ/_vcd	r = .84*	r = .54*	r = .76*	r = .85*		r = .84*	r = .80*	r = .87*
/o/_vls	r = .89*	r = .71*	r = .86*	r = .86*	r = .84*		r = .86*	r = .89*
/o/_vcd	r = .85*	r = .70*	r = .85*	r = .82*	r = .80*	r = .86*		r = .83*
/ɛ/_vls	r = .86*	r = .62*	r = .78*	r = .84*	r = .87*	r = .89*	r = .83*	

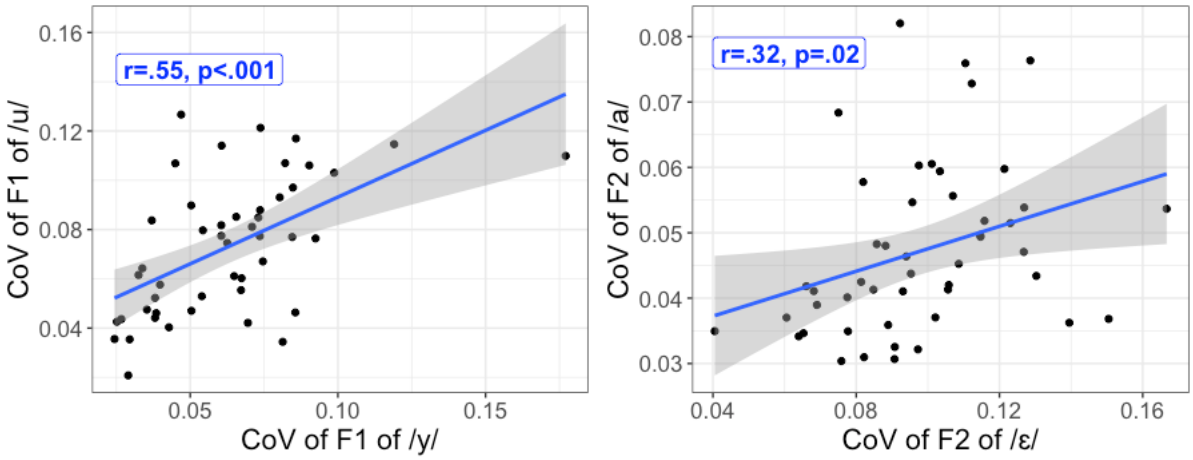


Figure 2.26 Significant Pearson correlations between the Coefficients of Variation of F1 in the vowel pair /y-u/ (left) and between the Coefficients of Variation of F2 in the vowel pair /a-ɛ/ (right).

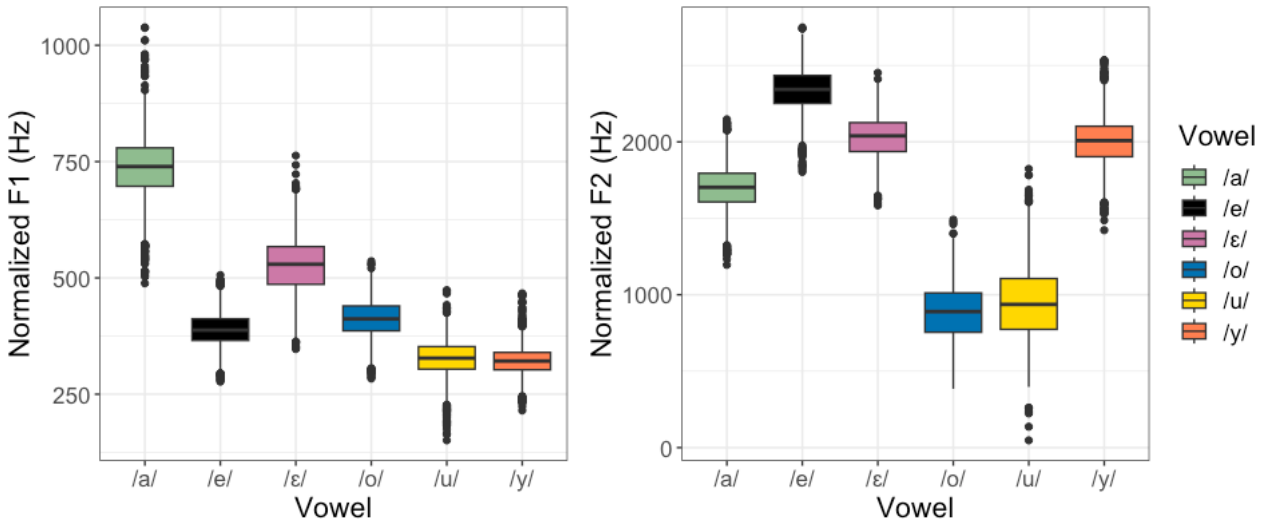


Figure 2.27 Boxplots of normalized F1 and F2 values of the four main vowels of interest in this study (/e, ɛ, o, u/) as well as two additional vowels (/a/ and /y/) whose variability is analyzed in this section.

No other significant correlation was observed between the CoV of F2 of /a/ and other vowels, given the greater differences in F2 of /a/ and the latter. Also unsurprisingly, no correlation was observed between the CoV of F1 of /a/ and any other vowel, as /a/ was the only low vowel present in the analyzed data. A weaker correlation was observed between CoV of F1 for the vowel pair /y-o/ ($r = .33$, $p = .02$), possibly due to both proximity in F1 values and similar lip rounding. No other correlation was found for CoV of F1 of /y/ nor for the CoV of F2 of /y/ with any other vowel. All correlation coefficients are given in Table 2.7 The strong correlation between the CoVs of F1 in the vowel pair /u-y/ mirrors the strong correlation between the CoVs of F1 of the vowel pair /e-o/, as shown in Figure 2.19. Similarly, the weaker correlation observed for CoV of F2 of the pair /a-ɛ/ is of a similar magnitude as those observed in Figure 2.20 for the vowel pairs /u-o/, /e-o/, /ɛ-o/ and /ɛ-e/.

Table 2.7 Pearson correlations between Coefficients of Variation of F1 or F2 of the four main vowels in the production experiment (/e, ɛ, o, u/) and the additional vowels /y/ and /a/. Significant correlations are indicated with one (p = .02) or two (p < .001) asterisks

Coefficient of Variation of F1				
	/e/	**/ɛ/**	**/o/**	**/u/**
/y/	r = .10	r = .03	**r = .33***	**r = .55****
/a/	r = .08	r = .03	r = .23	r = .09
Coefficient of Variation of F2				
	/e/	**/ɛ/**	**/o/**	**/u/**
/y/	r = -.01	r = .19	r = .11	r = -.01
/a/	r = .07	**r = .33***	r = .11	r = .06

As predicted, correlations in intraindividual variability of a certain acoustic measure are present when the two categories being compared share both a common articulatory gesture and similar resulting acoustic values. While the results in this chapter offer convincing evidence of this conclusion, some of the observed correlations, such as those between the CoVs of F2 of /e/ and /o/, as shown in Figure 2.20, suggest that it is not always possible to pinpoint an exact acoustic property or articulatory configuration that would explain the correlation in variability. However, it seems clear that variability is, for the most part, specific to distinct articulatory configurations which yield similar acoustic consequences. Similar results were reported by Whalen et al. (2018) for English vowels and Harper (2021) for English consonants. Given this relationship, it can be proposed that variability is, in a way, independent of phonological categories, or at the very least, as Harper (2021) proposes, is present only at the subsegmental or gestural level. In other words, if multiple segments share the same acoustic consequence and/or articulatory configuration, their rates of variability should be comparable, even if the segments themselves belong to distinct categories. It should be noted, however, that a change in the position of a single articulator can affect multiple acoustic parameters and a one-to-one correspondence between a specific articulator and a specific acoustic dimension is not plausible (Iskarous et al. 2011). The interpretation of the results in this chapter is that variability is

correlated across similar acoustic ranges that correspond to *general* articulatory configurations, for example, overall tongue shape rather than the configuration of a specific point on the tongue.

The different magnitudes of the correlations observed in this chapter bolster this interpretation. The strongest correlations were observed for similar rates of variability in f0 (all $r > .54$), voicing proportion (all $r > .64$) and F1 for the vowel pairs /o-e/ ($r = .83$) and /u-y/ ($r = .55$). The acoustic measures of f0 and voicing proportion are nearly entirely dependent on the glottal state, that is the precise configuration of the vocal folds. Although similar acoustic results for f0, for example, can be achieved by means of distinct articulatory strategies (e.g. a lower f0 can be due to slacker or shorter vocal folds), the general region of articulatory interest is identical. Again, speaker's average values for f0 and voicing proportion were highly similar across distinct segments, indicating that intraindividual variability is likely not dependent on highly specific articulatory configurations, but rather the resulting acoustic properties that stem from the same general articulatory region and configuration. The F1 values of the two high vowel pairs for which F1 variability was highly correlated (/o-e/ and /u-y/) were nearly identical, as seen in Figure 2.27 and presented a much stronger correlation than the remaining vowel pairs for which intraindividual variability was correlated, such as F2 of /u-o/ or F1 of /o-y/. Although the latter vowel pairs also shared similar formant ranges, they are crucially not identical and likely correspond to adjacent but semi-distinct regions in a speaker's vowel space, and thus slight differences in rates of intraindividual variability are expected.

In conclusion, the proposal that variability is correlated across specific acoustic ranges, regardless of the phonological segment measured, describes the tendencies observed in this chapter very well. There remains considerable interindividual variability and these correlations are not necessarily robust across all individuals. It is likely that each speaker has a default degree of variability for each acoustic range and/or articulatory configuration, but this variability is not entirely outside of their control. Additional factors, such as the laboratory setting or the relative unnaturalness of a forced-reading task, could have induced more careful speech or conscious use of articulators in some participants, thus altering their underlying variability patterns. Section 2.4 compares acoustic data of ten speakers across two recording sessions, precisely to better understand the consistency of intraindividual variability.

2.4 Inter-session variability (Production A and Production B)

As explained in Section 1.5, ten subjects participated in an initial production experiment (Production A) that took place between three to five months before the combined production-perception experiment (Production B). The target words, frame sentence and recording conditions were identical to those in Production B, as described in Section 2.1.2. This section presents the comparison of participants' productions across these two sessions to address Research question 1.2. Of interest are the comparison of mean acoustic values for each of the segments analyzed as well as rates of intraindividual variability, measured by the Coefficient of Variation (CoV). The acoustic analysis for Production A was identical to that described in Section 2.1.4 and Section 2.1.5. Namely, the following measurements were taken: post-release aspiration duration and voicing proportion for voiced and voiceless stops, mean values of F1 and F2 averaged over the steady state portion of the four vowels /e/, /ɛ/, /o/ and /u/ and the first four spectral moments for the fricatives /s/ and /ʃ/. Linear mixed-effect models were built using the package *lmerTest* (Kuznetsova et al. 2017) following the same procedure as described in Section 2.1.6. Specifically, all resulting models analyzed Production Session (**A** or B) as the fixed effect with a by-subject random intercept and slope for the variable Session. Additional fixed effects and interactions included for each phonetic category are explained in the following sections.

2.4.1 Voiced and voiceless stops

Overall, voicing proportion values of voiced and voiceless stops between Production A and Production B were similar for most participants, as shown in Figure 2.28. While individual differences are apparent, such as a higher rate of carryover voicing in voiceless stops or wane voicing in voiced stops in Production B for speakers 7 and 10, respectively, most subjects' patterns remained quite similar. To analyze group patterns, two mixed-effect models with proportion voicing as the dependent variable and Production Session as the predictor were carried out separately for voiced and voiceless stops. As the interaction of Place and Production Session was not significant and did not improve model fit and the objective here is to analyze

general intraindividual variability patterns, neither the interaction nor the predictor Place were included. These models corroborated a lack of significant difference between production sessions at the group level (**Voiced**: β = -0.94, t = -.92, p = .38; **Voiceless**: β = -5.1, t = -1.9, p = .1).

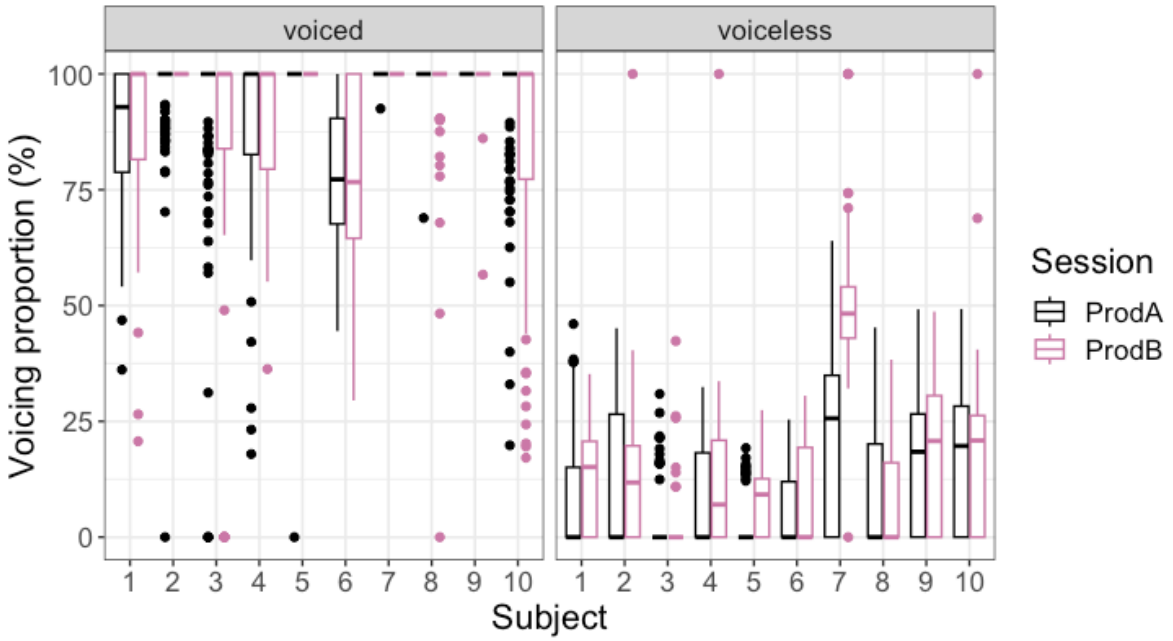


Figure 2.28 Voicing proportion for voiced and voiceless stops across both recording sessions (Production A and Production B) across all ten speakers.

Similarly, Figure 2.29 presents values of post-release aspiration duration in voiceless stops across both production sessions. The corresponding mixed-effect model confirmed no group-level significant differences between the two recording times (β = -.82, t = -0.67, p = .52).

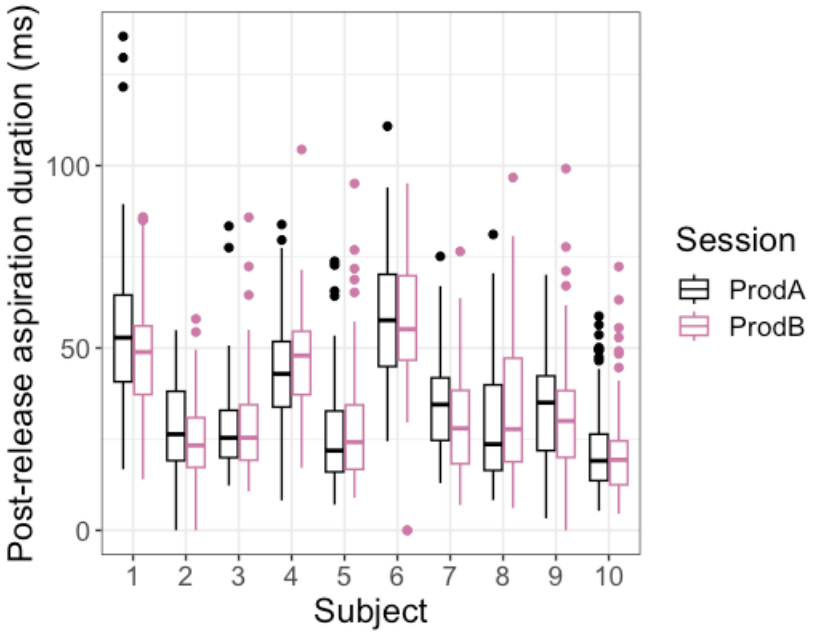


Figure 2.29 Post-release aspiration duration for voiceless stops in both recording sessions (Production A and Production B) across all ten speakers.

Variability was quantified by means of the Coefficient of Variation, calculated separately for post-release aspiration duration in voiceless stops and voicing proportion in voiced stops. In addition to similar mean acoustic values across production experiments, generally individuals differed little in the amount of variability present in their realization of voiced and voiceless stops, as seen in Figure 2.30. Slightly more inter-session variability was observed for voicing proportion, but at the group level, neither inter-session difference was statistically significant, as revealed by the mixed-effect models with CoV as the dependent variable and Production Session as the predictor (**Post-release aspiration**: β = -0.04, t = 1.8, p = .11; **Voicing Proportion**: β = -0.03, t = 1.5, p = .16). This indicates that the degree of intraindividual variability for a given acoustic property is a relatively stable characteristic of a speaker. If this inter-session stability of variability is observed across different acoustic properties, it would strongly suggest that variability is a robust aspect of one's speech production system.

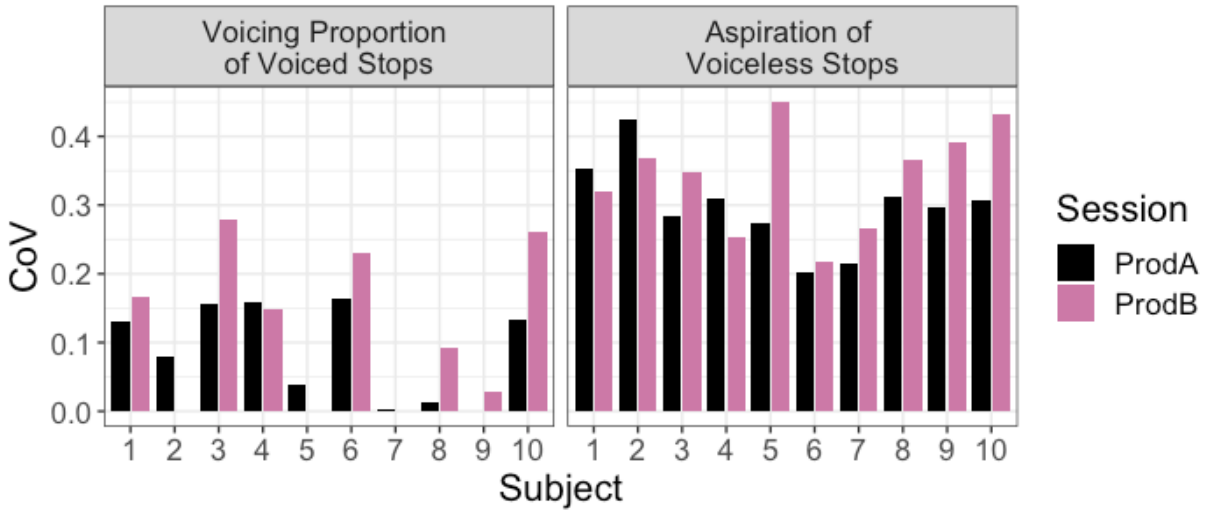


Figure 2.30 Coefficients of Variation (CoV) of voicing proportion (left) and post-release aspiration duration (right) for voiced and voiceless stops, respectively, across the ten speakers who participated in both production sessions (A and B).

2.4.2 Vowels

Figure 2.31 presents average normalized F1 and F2 values for each speaker across both production sessions. Each vowel label represents one speaker's mean production and the connecting lines represent the distance between vowel centroids. The normalization procedure was the same as described in Section 2.2.2. Two mixed-effects linear regressions were carried out to determine if F1 and F2 were significantly different between production sessions. Both models were built with the fixed effects Session, Vowel and their interaction and a by-subject random intercept and slope for Session. The dependent variables were normalized formant values. Post-hoc analyses revealed few significant differences. F1 was overall slightly higher for /u/ in Production B (β = -12, t = -26, p = .009) than in Production A and F2 was slightly higher for /o/ in Production B than in Production A (β = -44.8, t = -3.1, p = .001) (Table A14 and Table A15). Given that these differences are quite small, it can be claimed that participants' vowel distributions remain highly similar across recording sessions, but their mean values naturally differ given the flexibility of vowel targets which correspond to a general acoustic space and not to specific formant values. Given that the CoV values observed in Production B, as shown in Figure

2.17, were greatest for the back vowels /o/ and /u/, it is unsurprising that the group mean is slightly different across sessions.

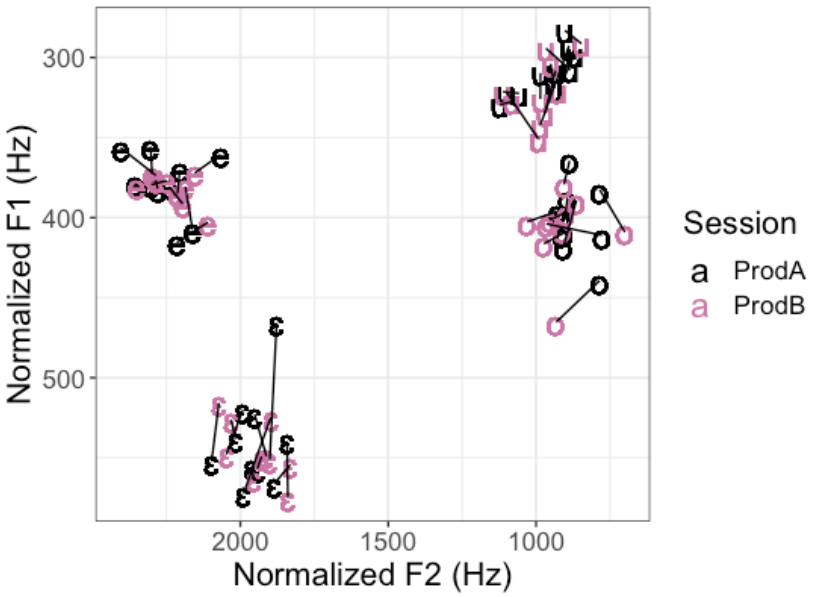


Figure 2.31 Normalized mean F1 and F2 values across production sessions A and B. The vowel symbols connected by a line correspond to those from the same speaker.

Euclidean distances between non-normalized vowel means across sessions were calculated to quantify the amount of change between vowel distributions, shown in Figure 2.32. A mixed-effects model on the Euclidean distances with Vowel as a fixed effect and by-subject random intercepts revealed no significant differences across vowels, indicating that the between-session differences are comparable across vowels (see Table A16). In general, this seems to indicate that no single vowel is more variable for all speakers across sessions.

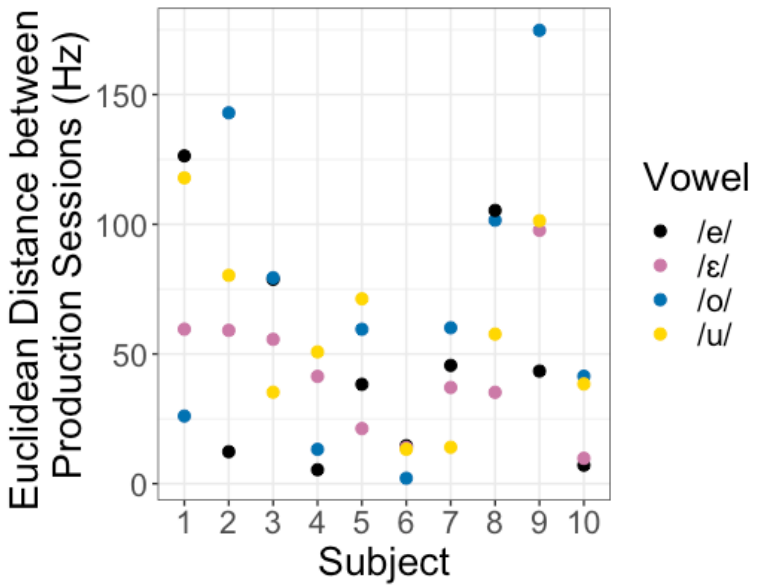


Figure 2.32 Euclidean distances, indicating the change in mean vocalic formant values, between Production A and Production B.

Coefficients of Variation (CoV) were calculated for F1 and F2 in order to compare the degree of variability between the two production sessions, presented in Figure 2.33. As observed in Section 2.2.2.2, back vowels had a larger degree of variability overall, coinciding with observations in other languages such as English, Parisian French, German (Strange et al. 2007) and Catalan (Recasens & Espinosa 2009). Two separate mixed-effect linear regressions, one for each formant dimension, with fixed effects of Session, Vowel and their interaction were carried out. No significant differences were observed in CoV values across the two production sessions, indicating that speakers' degrees of variability for an individual segment are stable (see Table A17 and Table A18). It is particularly interesting that participants' mean acoustic values presented substantially greater differences than their rates of variability. Again, this indicates that while formant values themselves may vary slightly between recording sessions, the overall vowel spread is quite stable.

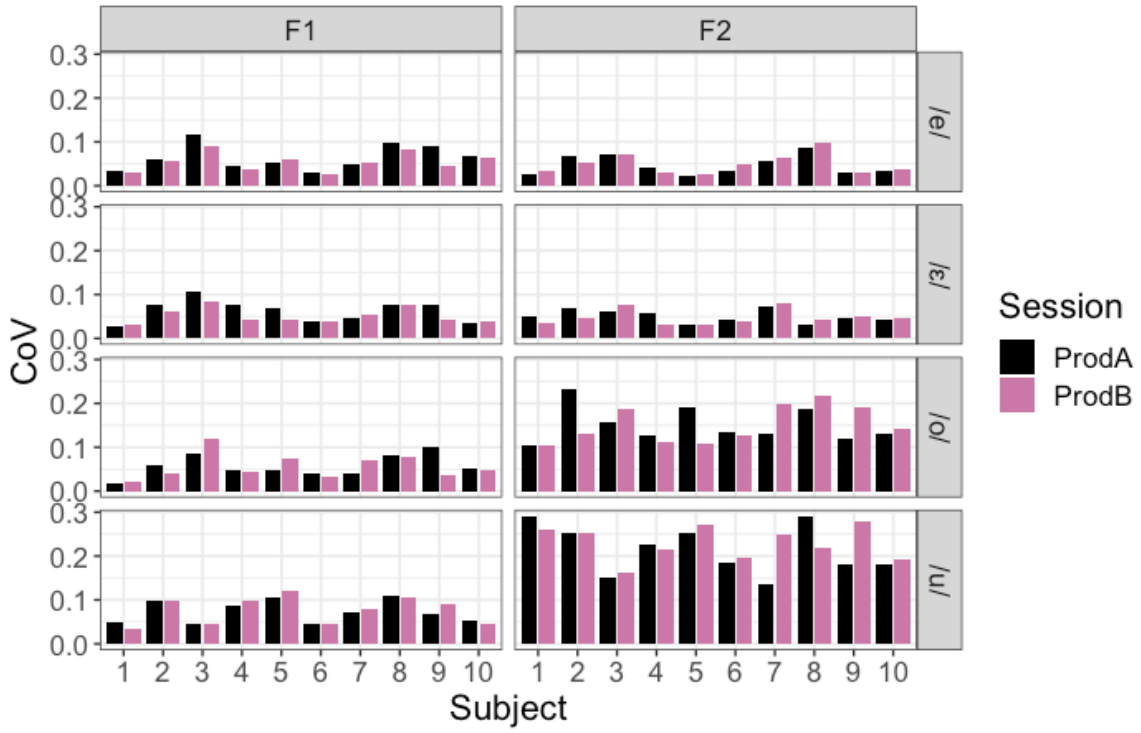


Figure 2.33 Coefficients of Variation (CoV) of F1 and F2 for the four vowels of interest across the ten speakers who participated in both production sessions (A and B).

2.4.3 Fricatives

The four spectral moments of the two coronal fricatives were extracted in the same way as for Production B, namely averaged over the portion of the fricative excluding the first and last 10%, as described in Section 2.2.2 (DiCanio 2013). Figure 2.34 presents the normalized values from the 10 participants in both production experiments. Mixed-effects models were built with the fixed effects Session, Place and their interaction and by-subject random intercepts and slopes for Session (see Table A19). Post-hoc analyses revealed significantly higher COG values for /ʃ/ (β = -205, z = -3.4, p = .003) and /s/ in Production B (β = -186, z = 2.9, p = .007). Importantly, both fricatives presented higher COG values in Production B, indicating an overall increase in frequency rather than one fricative being affected more than the other. No significant differences were observed across sessions for the remaining spectral measures (p > .2). It is unclear what may have caused this slight increase in COG. Given no similar effects were observed for stops or

vowels, it is unlikely to be because participants had already produced the list of stimuli in Production A. Natural fluctuations due to intraindividual variability are likely the cause.

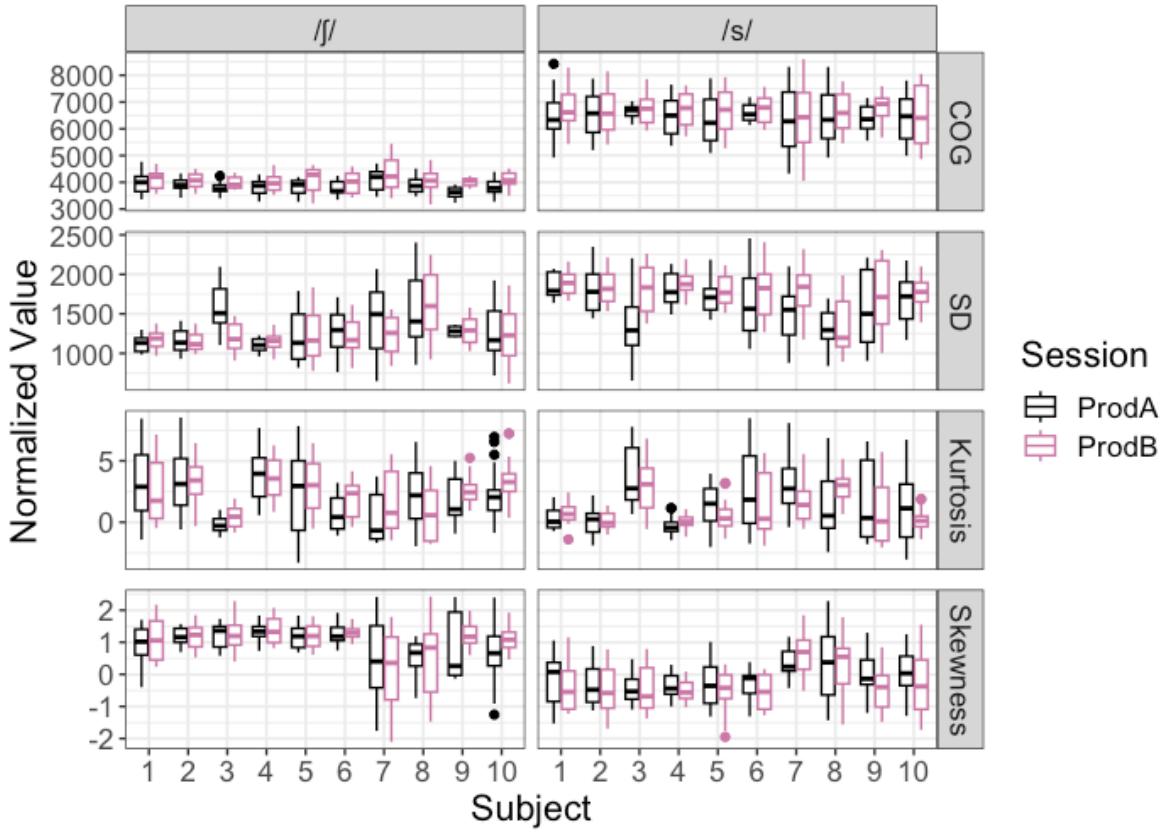


Figure 2.34 Boxplots of the first four spectral moments of the dental (/s/) and alveopalatal (/ʃ/) fricative across production sessions.

As in Production B, given the overwhelming significance of COG in distinguishing the two fricative categories, in comparison to the other spectral measures reported, variability was analyzed exclusively for COG by means of CoV, as shown in Figure 2.35. Highly similar rates of variability were observed for both fricatives in most participants, with only a few exceptions. Post-hoc analyses on the mixed-effect models with Session, Place and their interaction as fixed effects confirmed that the CoV for COG did not significantly differ between Production A and Production B for either of the two fricatives (**Fricative/s/**: β = -0.009 t = .9 p = .38; **Fricative/ʃ/**: β = -0.01 t = -1.5 p = .16, see Table A20). This again suggests that the magnitude of intraindividual variability for a given acoustic segment is a unique and stable property of an individual's speech.

It should be noted that the degree of variability of /ʃ/ for Participant 5 and /s/ for Participant 10 were found to be considerably greater in Production B. It is unclear what conditioned the increase in variability for these participants. While slight fluctuations in rates of variability are to be expected, it is possible that an external factor affected the speech of fricative production of these two individuals during Production B. Without additional data, further speculation is unwarranted.

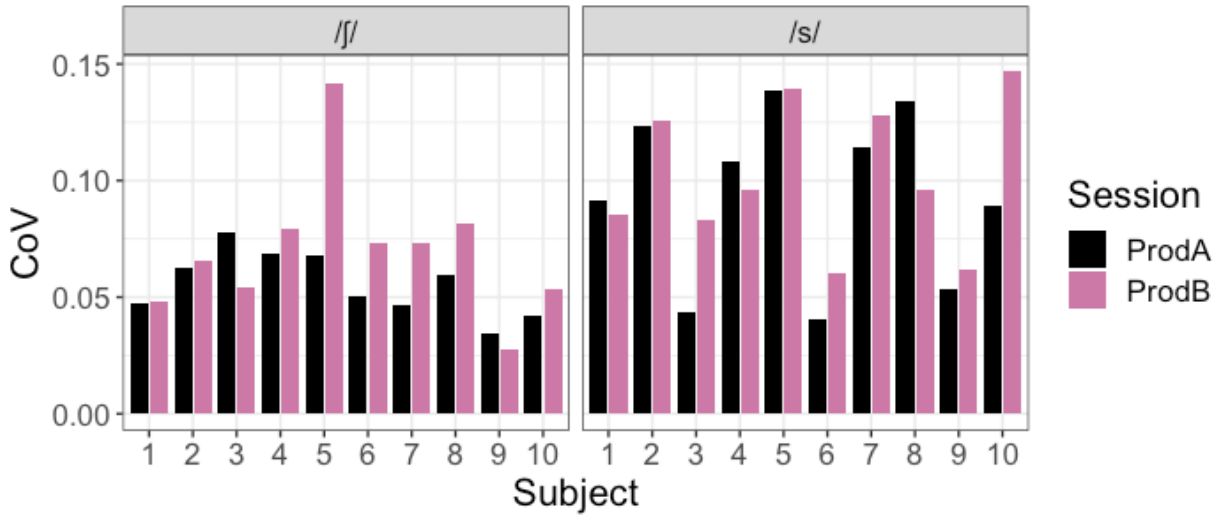


Figure 2.35 Coefficients of Variation (CoV) of COG for the dental (/s/) and alveopalatal (/ʃ/) fricative across the ten speakers who participated in both production sessions (A and B).

2.4.4 Summary

Overall, little change across production sessions was observed, suggesting speakers' average acoustic ranges appear to be quite stable, as reported by Heald & Nusbaum (2015). F1 of /u/ and F2 of /o/ were slightly higher in Production B than in Production A and COG values of both fricatives were higher in Production B than in Production A. These differences were relatively small and may be attributed to random fluctuations due to the different time of day at which participants carried out the recordings across the two experimental sessions. Indeed, such an effect has been observed previously by Heald & Nusbaum (2015). Slight variation in f0 could also have influenced both formant values and fricative spectral characteristics. That said, it is essential to emphasize the difficulty in comparing mean values across sessions in the presence

of intraindividual variability. Indeed, the results from Production B alone confirmed that individuals often vary considerably from one token to the next. All things considered, variability across sessions can be understood as forming part of the larger picture of an individual's rate of variability. In other words, if all experimental conditions are the same as they were in the current study, differences between sessions are essentially equal to differences between tokens within a single session.

Even more crucial to the central research questions of this dissertation is the lack of significant differences between CoV across production sessions. In other words, in addition to general acoustic ranges, the rate of intraindividual variability for specific segments appears to be a stable characteristic of an individual's speech. This suggests that variability is not a random characteristic, but rather a structured element built into an individual's phonological grammar in some way. This will be further discussed in Chapter 5.

2.5 Conclusion

Intraindividual variability was consistently observed in the speech of all participants and for all phonological segments tested. The degree of variability differed considerably across participants, suggesting that the tolerance level for variability within one's own speech is a unique characteristic of the speaker. Rates of variability were found to be correlated across segments that shared acoustic ranges and therefore similar articulatory configurations. While this relationship was not observed for every individual, its consistent significance across a diverse range of acoustic properties involving distinct articulatory mechanisms offers strong evidence that variability is neither a characteristic of an individual's overall speech nor specific to a precise phonological or phonetic segment. Instead, intraindividual variability is structured and affects different articulatory configurations and their resulting acoustic consequences disproportionately and in distinct ways for each individual. These rates of variability are a consistent property of an individual, as observed in the lack of significant changes across recordings carried out months apart, thus reinforcing the idea that variability is representative of an underlying, predictable structure. It is already clear that considering both inter- and

intraindividual differences in speech production can reveal distinct patterns that open the door to many directions of future research. The organization of such intraindividual variability in perception will be explored in Chapter 3.

3. Perception experiment

A two-alternative forced choice (2AFC) identification experiment was designed in OpenSesame (Mathôt et al. 2012) to test perception of the relevant acoustic measures. It is noted that some recent perception studies in phonetics have been conducted with a visual analogue scaling (VAS) task, rather than a 2AFC task, under the assumption that results from a VAS task disentangle gradient perception, that is the ability of listeners to pick up on fine-grained acoustic details without the need to categorize sounds binarily, and response inconsistency (Kapnoula et al. 2017; Apfelbaum et al. 2022; Honda et al. 2024a, among others). In a VAS task, participants are instructed to rate each token heard on a sliding scale between the two elements of a minimal pair, indicating to what degree they believe a heard token belongs to one of the two provided categories. Although one argument in favor of the VAS task is that it does not force gradient perceivers to categorize stimuli on a binary scale, as is the case in a 2AFC task, the VAS task runs into the opposite issue, namely that less gradient perceivers may artificially use the scale under the assumption that that is what is expected of them by the task itself. Additionally, the VAS task likely increases memory demands on participants, possibly leading to less attentiveness and quicker task-exhaustion. In sum, given that the advantage offered by the VAS task of better capturing gradient perception is not pertinent to the objectives of the current dissertation and to avoid participant fatigue in the combined production-perception experimental session, the 2AFC method was deemed more appropriate for the purposes of the current study.

3.1 Methodology

3.1.1 Participants

All participants in this perception experiment were the same as in the preceding production experiments and they completed the perception task immediately following the recording of Production B.

3.1.2 Stimuli

Three sets of stimuli were generated to investigate the three phonological and acoustic categories chosen for the current study: stop voicing, vowel pairs and voiceless coronal fricatives. All stimuli were chosen to be actual French words so as to increase the likelihood participants accessed real speech perception strategies. To build the stimuli, five repetitions of all words presented in Table 3.1 were first produced by a native Laurentian French female speaker and the clearest utterances were further manipulated across the relevant dimensions as detailed in the following sections. To ensure an accurate representation of this speaker's vocalic space, five repetitions of the following additional words were recorded: *goût* /gu/ 'taste', *cou* /ku/ 'neck', *fou* /fu/ 'crazy', *daté* /date/ 'dated', *rater* /ʀate/ 'to miss', *atout* /atu/ 'skill'. The inclusion of these additional words allowed for approximately the same amount of datapoints to be obtained for the model speaker as for all participants in Production B. In addition to allowing for a more accurate representation of the model speaker's formant ranges, the formant values of all tokens produced by the model speaker were used for normalization purposes upon analyzing participants' production-perception links, as will be explained in detail in Chapter 4.

Table 3.1 Stimuli used in the perception experiment for the three phonological categories tested.

STOP VOICING			
/b-p/		**/g-k/**	
/o/	**/ɛ/**	**/o/**	**/ɛ/**
les beaux /lebo/ 'the beautiful ones'	*la belle* /labɛl/ 'the beautiful one'	*l'égo* /lego/ 'the ego	*pagaie* /pagɛ/ 'oar'
les peaux /lepo/ 'the skins'	*la pelle* /lapɛl/ 'the shovel'	*l'écho* /leko/ 'the echo'	*paquet* /pakɛ/ 'package'
VOWELS			
/u-o/		**/e-ɛ/**	
tout /tu/ 'all'		*fée* /fe/ 'fairy'	
taux /to/ 'rate'		*fait* /fɛ/ 'done'	
FRICATIVES			
lasser /lɑse/ 'to bore'		*lâcher* /lɑʃe/ 'to release'	

3.1.2.1 Voiced and voiceless stops

As confirmed by the results in Section 2.2.1, many voiced stops are produced with incomplete closure voicing and often with post-release aspiration. In order to test the relevance of each of these cues in the voicing distinction, stimuli were manipulated across these two properties, resulting in tokens with varying degrees of partial closure voicing and post-release aspiration duration. Two places of articulation preceding two vowels were considered, for a total of four stop voicing pairs as presented in Table 3.1.

The first step in creating the stimuli was to extract the vowels following the voiced and voiceless stops in each pair and neutralize formant values in order to prevent listeners from using coarticulatory information to determine the voicing of the original segment. This was achieved by means of a Praat script that produces a formant continuum between two reference vowels (Winn 2014a; Winn & Litovsky 2015; Boersma & Weenink 2022). First, each vowel was inverse-filtered using the linear predictive coding (LPC) decomposition procedure in Praat to obtain the spectral envelope of the speech sound without formant peaks. Five interpolated steps between the formant contours of the base post-voiceless vowel and the base post-voiced vowel were calculated for the first third of the vowel, where formant transitions differ between voiced and voiceless stops. As higher frequencies are lost in the LPC procedure, the original vowel token was high-pass filtered at 3500 Hz and added back to the low-pass filtered manipulated vowel with a gradual crossover of 500 Hz. The third step of this generated continuum was retained and used as the base vowel for the following additional manipulations. This third step was chosen as it corresponds to the most neutral vowel between the two voicing categories.

Following this, three steps of aspiration were determined for both places of articulation, encompassing average values of post-release aspiration in voiceless stops as well as in partially voiced stops, which is often shorter. These three steps of aspiration were created by shortening the aspiration from the base voiceless stimulus. To ensure the initial burst remained constant, aspiration was cut off from the end, closest to the onset of the following vowel. To ensure participants could not make use of f0 as a secondary cue to voicing, the f0 of the first third of the post-release vowel was held constant at an intermediate value between average f0 values for the voiced and voiceless stops produced by the base speaker.

As shown in Chapter 2, a large portion of voiced stops were realized with wane voicing, that is, voicing that begins at the start of the closure and disappears before the release. Additionally, no tokens were observed in which the start of the closure of a voiced stops was voiceless and voicing subsequently initiated in the middle of the closure. Thus, to test participants' sensitivity to these observed voicing patterns, four durations of closure voicing were generated as follows. First, the final 10 ms of a voiced token produced with incomplete closure voicing was extracted to ensure that voicing disappeared gradually before the release, as is observed in natural stimuli. Silence was then appended to the end of this voicing duration to create Step 1. Subsequent steps were created by adding distinct durations of complete closure voicing to the onset of Step 1. The closure duration was held constant across all stimuli. The voicing steps were then added to the previously created aspiration steps and these newly created stimuli were inserted into the original words. The duration of the aspiration and closure voicing steps are shown in Table 3.2. All 12 (4 voicing x 3 aspiration) combinations were included in the experiment. An example oscillogram and spectrogram of Step 3 voicing (flanked by solid red lines) and Step 3 aspiration (flanked by dashed blue lines) for the stimulus *la b-pelle* (only /ab-pɛ/ is shown) is seen in Figure 3.1. The dashed red line indicates the start of the final 10 ms of voicing that was extracted from a token produced with incomplete closure voicing. This stimulus is therefore quite similar to the naturally observed voicing pattern (C) as shown in Figure 2.3.

Table 3.2 Duration of steps of closure voicing and post-release aspiration used to build the stop voicing continuum in the perception experiment.

	Closure voicing (Closure duration = 105 ms)				**Post-release aspiration**		
	Step 1V	Step 2V	Step 3V	Step 4V	Step 1A	Step 2A	Step 3A
/b-p/	10 ms	42 ms	75 ms	105 ms	0 ms	18 ms	35 ms
/g-k/	10 ms	42 ms	75 ms	105 ms	0 ms	22 ms	45 ms

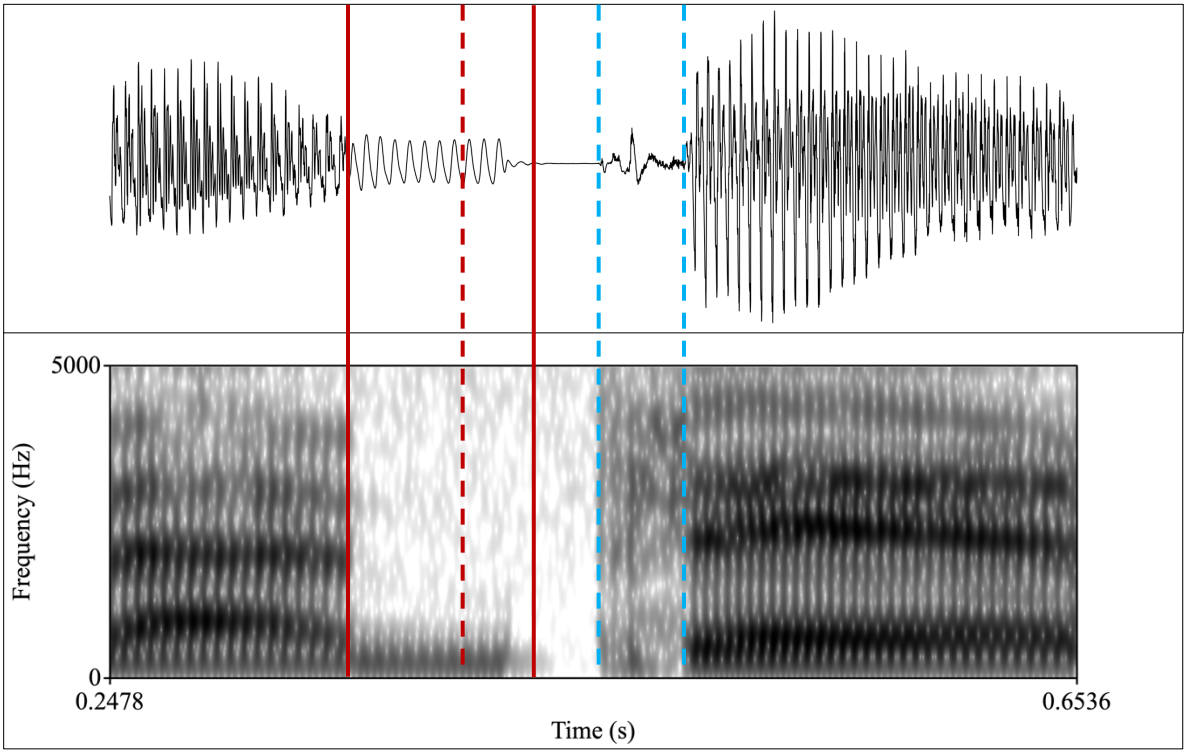


Figure 3.1 Oscillogram and spectrogram of an example stimulus *la belle* (only /ab-pɛ/ shown). The solid red lines encompass 75 ms of closure voicing, corresponding to Step 3V. The dashed red line indicates the start of the 10ms of voicing wane extracted from a token in which voicing disappeared during the closure. The dashed blue lines encompass 35 ms of aspiration, corresponding to Step 3A.

3.1.2.2 Vowels

The two pairs of target vowels were the same as those in the production experiments, as presented in Table 3.1, namely /e-ɛ/ and /u-o/. The onset consonants (/t/ and /f/) were selected specifically to differ from the onset consonants of the stop voicing stimuli. Five interpolated steps of F1 and F2 were created in Praat for each of the pairs following the same procedure as described in the previous section for the vowel in the stop voicing continua (Winn 2014a; Winn & Litovsky 2015; Boersma & Weenink 2022). Figure 3.2 presents the resulting formant values for the /e-ɛ/ continuum and Figure 3.3 the formant values for the /u-o/ continuum.

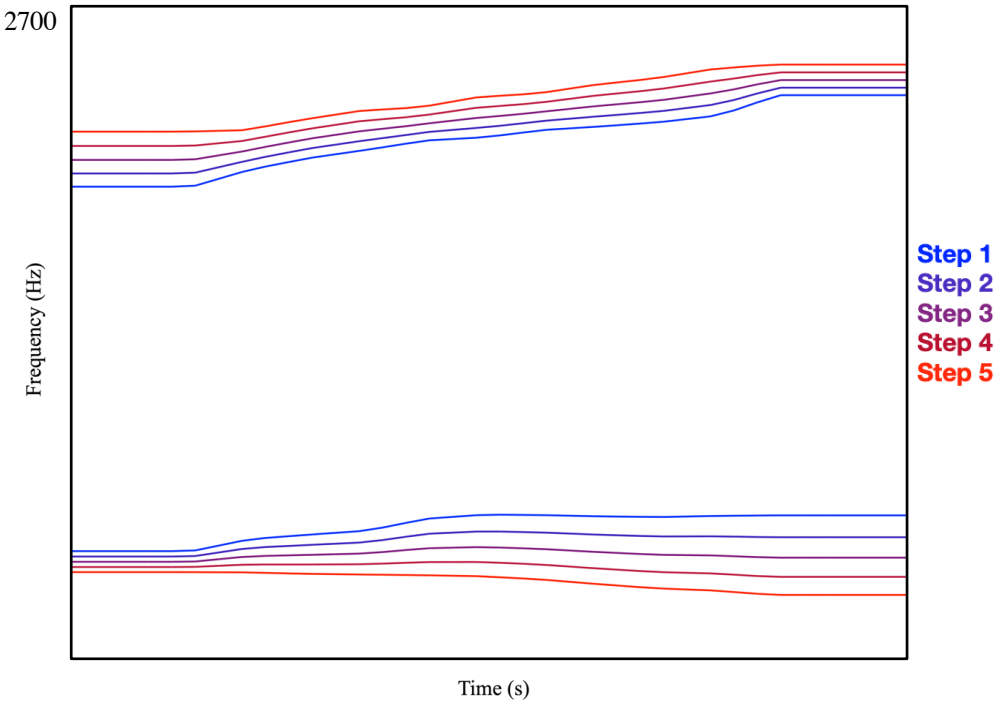


Figure 3.2 F1 and F2 values of the five steps of the formant continuum between /e/ and /ɛ/ following /f/. Step 1 (in blue) corresponds to /ɛ/ and Step 5 (in red) corresponds to /e/.

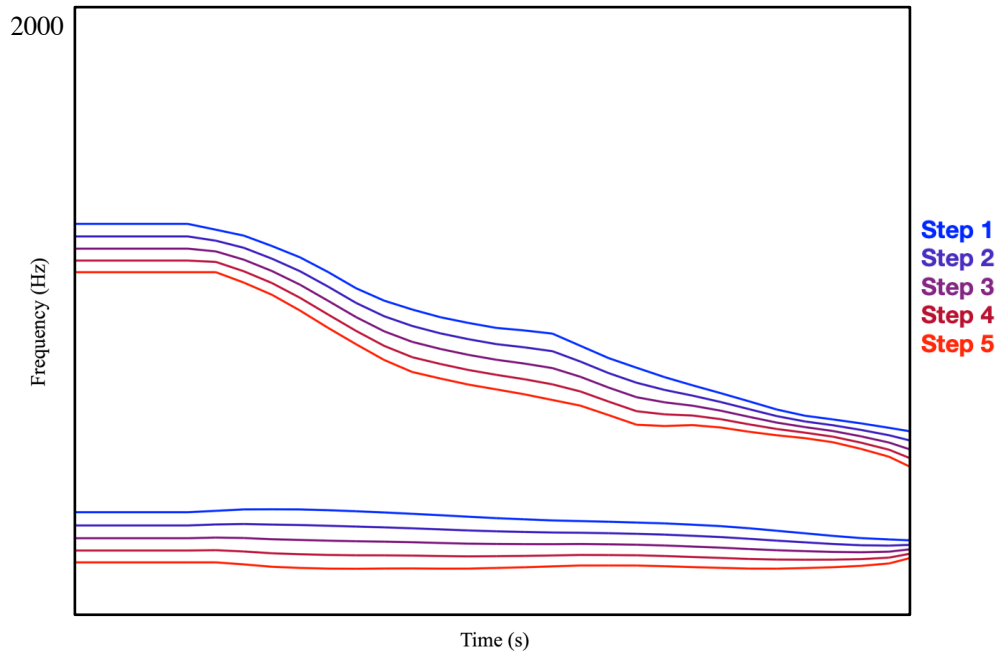


Figure 3.3 F1 and F2 values of the five steps of the formant continuum between /o/ and /u/ following /t/. Step 1 (in blue) corresponds to /o/ and Step 5 (in red) corresponds to /u/.

3.1.2.3 Fricatives

The minimal pair of *lasser* /lɑse/ ‘to bore’ and *lâcher* /lɑʃe/ ‘to release’ was chosen to test the perception of place of articulation in fricatives. Given the highly salient role of center of gravity (COG) previously reported in Laurentian French (Toda 2009; Salinas Marchant 2025), as well as the overwhelming importance of this cue in the results from Production A, as shown in Section 2.4, COG was chosen as the parameter to be manipulated. Using a Praat script, five interpolated steps were created between average COG values of /s/ and /ʃ/ produced by the example speaker (Winn et al. 2013; Winn 2014b; Boersma & Weenink 2022). White noise was generated at three distinct peak frequencies (1000 Hz, 4500 Hz et 8500 Hz) and the relative amplitudes between these peaks was manipulated across five steps ranging from -15 dB to 10 dB. These three peaks were then combined and an amplitude increase at onset and decrease at offset was applied to achieve a sound more closely resembling a natural fricative. Finally, each generated fricative was placed in between /lɑ/ and /e/ from the original stimulus of *lasser*, with 5 ms of crossfading on both sides to ensure smooth transitions between the segments. Using a synthetic fricative over a natural production was considered more adapted to the current study, as it is easier to hold all but one parameter constant, as is often done in studies testing fricative perception (Chodroff & Wilson 2020; Harper 2021; Uluşahin et al. 2024). The values for all parameters were verified with the aggregated values taken from the tokens produced by the speakers in Production A. Figure 3.4 presents the spectra of the five-step stimuli fricative continuum.

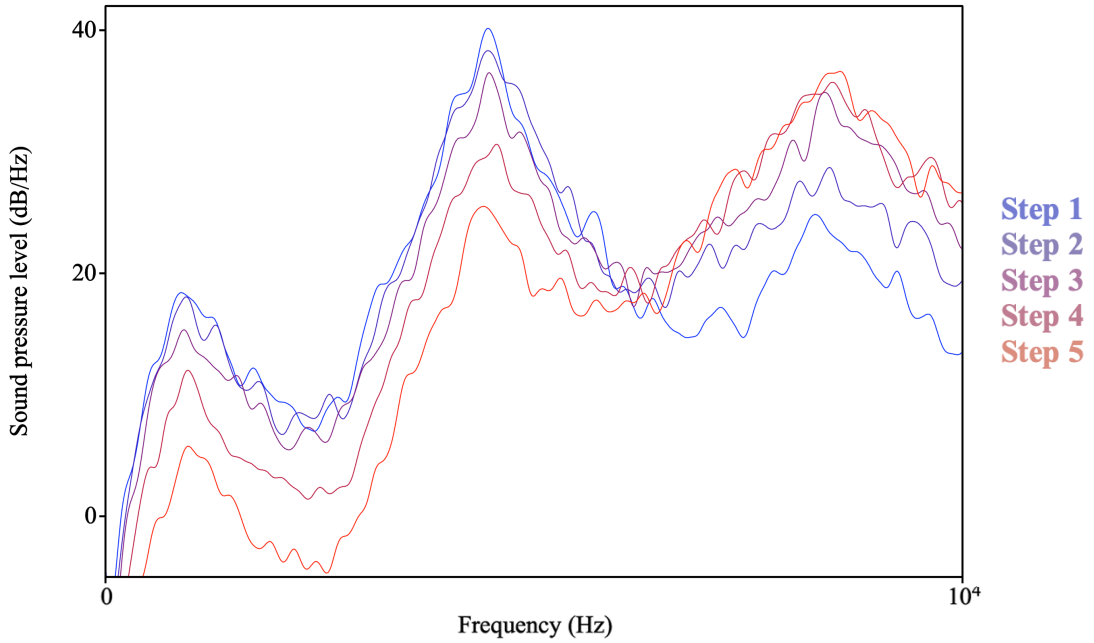


Figure 3.4 Spectra of the five steps of the fricative continuum. Step 1 (in blue) corresponds to /ʃ/ and Step 5 (in red) corresponds to /s/.

3.1.3 Procedure and analysis of response data

The 2AFC experiment was designed and carried out in OpenSesame (Mathôt et al. 2012). Choices were presented orthographically on a screen in two separate rectangles, as shown in Figure 3.5. Two blocks were built and each was presented twice. The composition of the blocks is shown in Table 3.3. Within each block, each token was repeated four times. The order of tokens was randomized and presented in the same randomized order for all participants and for both block repetitions. Overall, each participant heard the same token eight times for a total of 504 responses.

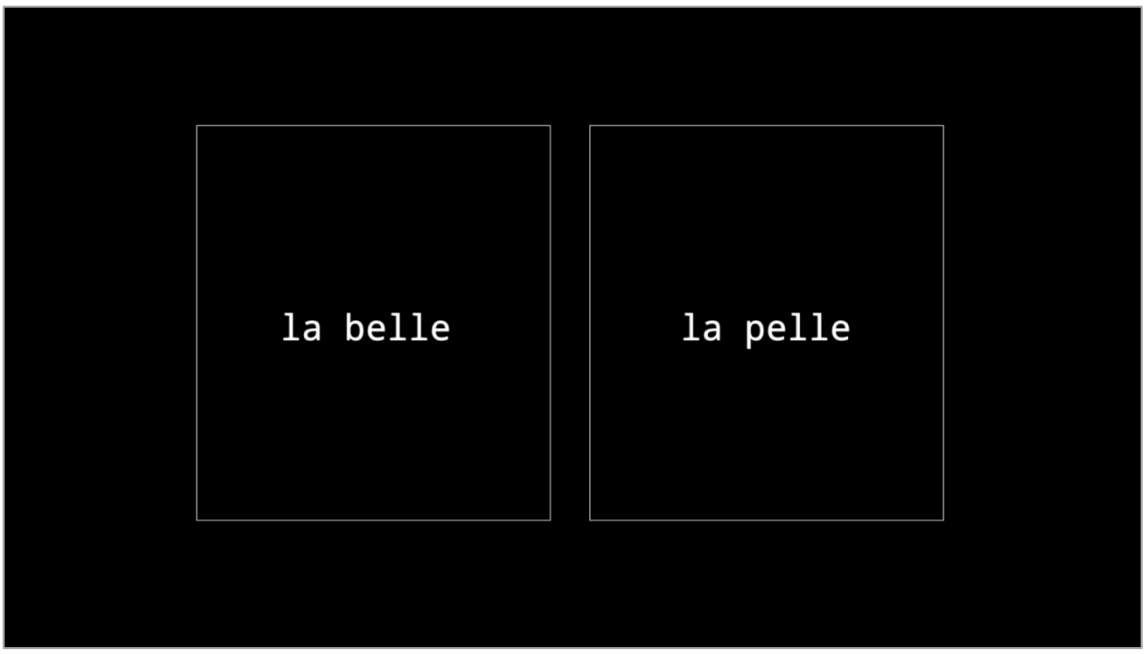


Figure 3.5 Example presentation of stimuli in the two-alternative forced choice perception experiment conducted with OpenSesame (Mathôt et al. 2012).

Table 3.3 Composition of two blocks in the two-alternative forced choice perception experiment. Within each block, each token was heard four times. Each block was repeated twice for a total of eight repetitions per token. Overall, 504 participant responses were collected.

Block 1		**Block 2**	
Continuum	**Number of stimuli**	**Continuum**	**Number of stimuli**
Vowel /u-o/	5	Vowel /e-ɛ/	5
Stop voicing /b-pɛ/	12	Stop voicing /b-po/	12
Stop voicing /g-kɛ/	12	Stop voicing /g-ko/	12
Fricative /s-ʃ/	5		
TOTAL:	**34**	**TOTAL:**	**29**

To neutralize possible handedness effects, the position of the two choices was switched in the second presentation of each block. That is, the words that appeared in the left box in the first block, appeared in the right box in the second block and vice versa. For each token, participants heard the stimulus and then immediately saw the two possible words on the screen. Upon clicking on the desired choice, the following trial began immediately. If participants clicked outside of the white rectangle surrounding each word, the same stimulus was played a second time. Participants completed the experiment on a MacBook Air, wore Sennheiser HD 280 pro headphones and used a mouse to click on each response. Prior to beginning, participants were instructed to choose the word they heard without overthinking and completed five practice tokens with words not used in the actual experiment before initiating. Responses with reaction times above 3000 ms or below 300 ms were excluded.

3.1.4 Analysis of variability

Previous studies have quantified the amount of variability in an individual's responses in a perception identification task by measuring boundary width (McAllister Byun & Tiede 2017; Park et al. 2019; Cheng et al. 2021). This is obtained by first fitting by-subject quasibinomial logistic functions over the count of one of the binary responses (here this means VOICED, /ɛ/, /o/ and /ʃ/). The predicted values are then extracted from each function and the value on the x-axis (continuum step) corresponding to 25% probability is subtracted from the value corresponding to 75% probability, as illustrated in Figure 3.6. A larger boundary width is indicative of greater indecisiveness when identifying ambiguous stimuli close to the theoretical categorical boundary, while a narrow boundary width indicates a much sharper slope and thus greater response consistency.

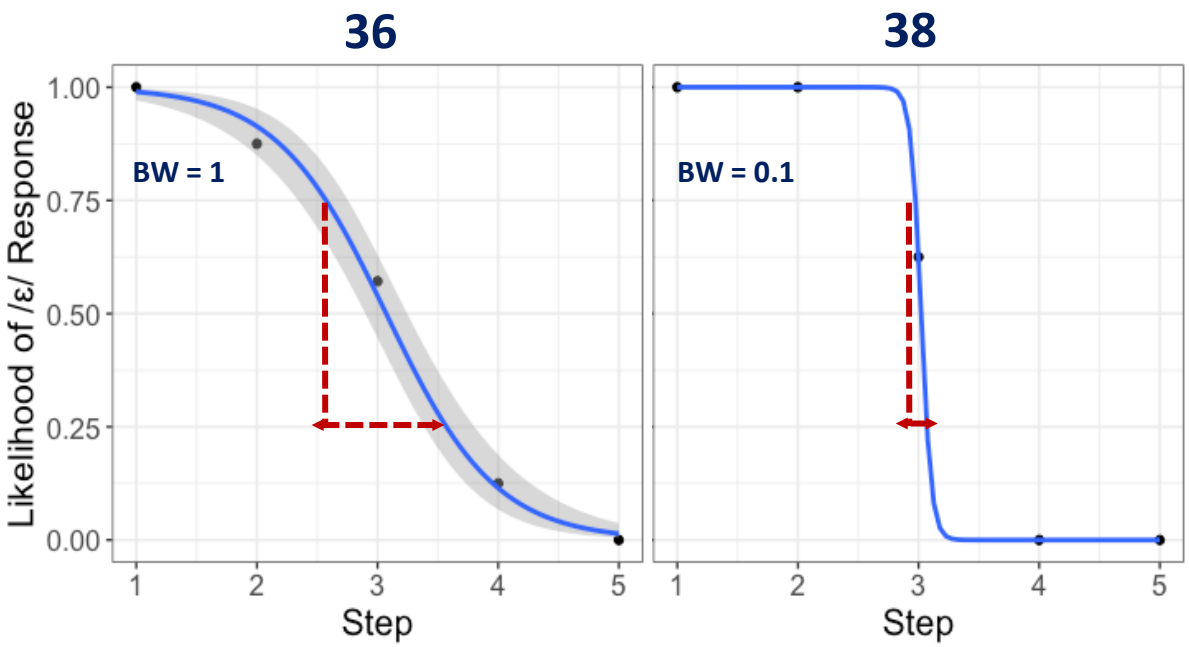


Figure 3.6 Visualization of the calculation of boundary width by means of a quasibinomial logistic function for a participant with a shallower slope (36) and a steep slope (38).

This measure is successful in quantifying the overall presence of variability but is less effective in analyzing variability in participants whose perceptual trajectories do not follow the expected S-curve often taken as an indicator of "expected" perception. The idea that perception is purely categorical for all individuals is nowadays largely contested (Apfelbaum et al. 2022), and thus excluding participants whose responses do not fully range from 0% to 100% would lead to a misrepresentation of actual perception. Given that the boundary width measure is best suited for continua that differ across a single acoustic dimension, it was not calculated for the stop voicing continuum in the current experiment, which varied across both voicing proportion and post-release aspiration duration. Despite these drawbacks, boundary widths were calculated for the vowel and fricative continua and will be used when analyzing the production-perception link in Section 4.2 and Section 4.3. However, in order to utilize the same measure across all continua here and to capture response (in)consistency across individual tokens, corresponding more closely to the type of intraindividual variability measured in the production experiment, an additional measure was used to quantify intraindividual variability, namely the Coefficient of

Unalikeability (CoU). The CoU is a way to quantify the amount of variability present in categorical data and is calculated using the following formula:

$$CoU = 1 - \sum_i p_i^2$$

In this formula, *i* = number of categories (here that is the binary response variable) and *p* = proportion of responses in each category (Kader & Perry 2007). The CoU can have a value between 0 (the same response for each of the eight repetitions) and 1 (a different response every time) and directly reflects the consistency of observations in a single category. Given that the response variable in the perception experiment is binary, the highest CoU possible is 0.5, which would indicate that each of the two responses was chosen equally. This coefficient has been used in previous linguistic studies to measure linguistic diversity in a population (Greenberg 1956) or consistency in an L2 noun gender labeling task (Shantz 2018). For data from the perception experiment in this study, CoU was calculated separately for each of the seven continua presented in Table 3.3. Below is an example calculation of the CoU for one speaker's responses for the /b-po/ continuum:

$$CoU = 1 - (\frac{3}{8})^2 + (\frac{5}{8})^2$$
$$CoU = 1 - 0.141 + 0.391$$
$$\boldsymbol{CoU = 0.468}$$

Additionally, a separate CoU value was calculated for larger groupings of the stop voicing continua by averaging together the individual CoUs. This yielded distinct values for the following additional categories: Bilabial stops (/b-pɛ/ & /b-po/), Velar stops (/g-kɛ/ & /g-ko/), and Stops (Bilabial and Velar). As shown in Figure 3.7, the CoU and boundary width measures were very strongly correlated, validating that both capture the same aspect of intraindividual variability. Given that a CoU value can be calculated for all continua and for all participants, regardless of the categoricity of their overall response slope, the CoU was chosen as the sole measure of intraindividual variability in perception.

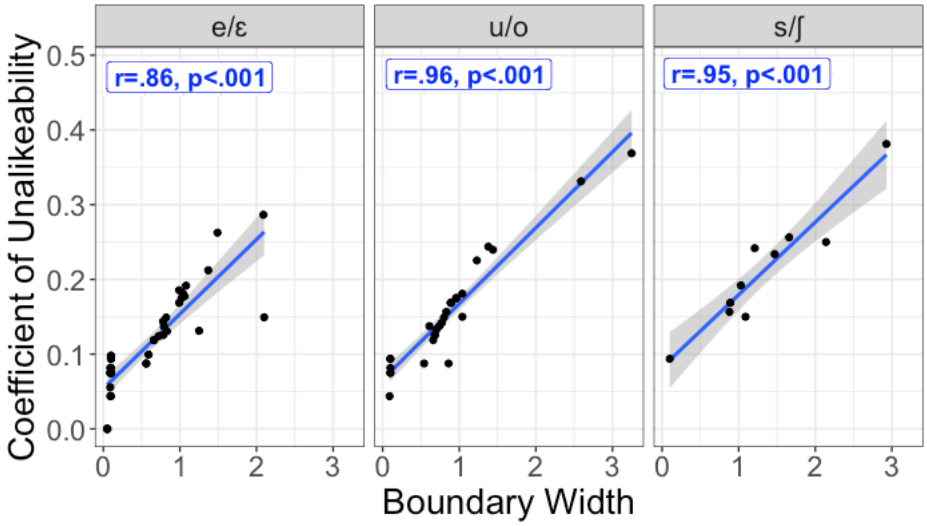


Figure 3.7 Significant Pearson correlations between individual values of boundary width and Coefficient of Unalikeability.

3.1.5 Statistical analysis

For the stop voicing continuum, given that the relative weights of post-release aspiration duration and closure voicing have not been previously analyzed in Laurentian French, it was deemed pertinent to analyze group-level tendencies by logistic regression. The dependent variable was the participant response coded as numeric (0 = voiceless and 1 = voiced), the predictors were the aspiration step (numeric), the closure voicing step (numeric), PLACE (**bilabial**, velar) and the three possible two-way interactions among these three predictors. By-participant random intercepts and slopes were included for both the aspiration and closure voicing steps. Group-level tendencies were not analyzed for the vowel and fricative continua, as such patterns are not relevant to the research questions of this dissertation. Relationships between degrees of intraindividual variability across distinct continua were analyzed by means of Pearson correlations of the CoU values. All statistical analyses were carried out in R (R Core Team 2022) and information on specific model structures are provided throughout the text.

3.2 Results

3.2.1 Voiced and voiceless stops

As seen in Figure 3.8 and confirmed by the logistic regression model (see Table A21), the group-level tendency is that the probability of a VOICED response increases as voicing duration increases (β = 1.8, z = 15 p < .001) and decreases as aspiration increases (β = -3, z = -24 p < .001). The interaction of aspiration and voicing was also significant, indicating that participants chose VOICED responses even with longer durations of aspiration if voicing duration increased concurrently (β = 0.15, z = 3.9 p < .001). Significantly more VOICED responses were recorded for the bilabial pairs (β = -3.1, z = -30 p < .001).

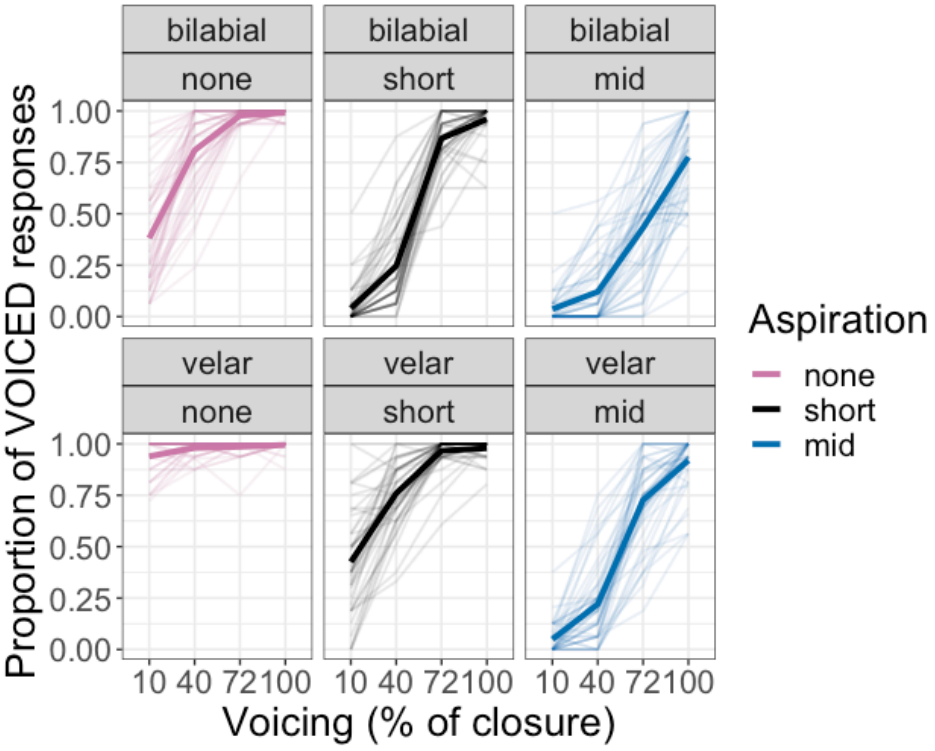


Figure 3.8 Proportion of VOICED responses for the stop voicing contrasts in the perception experiment. Thick, dark lines represent the mean across all participants, while pale lines show individual trajectories.

As an example, even when the closure is almost entirely unvoiced, a velar stop is still overwhelmingly perceived as voiced in the absence of aspiration. This coincides with tendencies observed in the production data where many fully voiced /g/ tokens were produced with short durations of post-release aspiration. In other words, aspiration alone is insufficient to identify a voiceless velar stop but the *lack of aspiration* is sufficient to identify a voiced velar stop, regardless of the presence of closure voicing. In completely or near-completely voiced bilabial stops, the presence of a longer duration of aspiration lowers the likelihood of the listener perceiving /b/, again mirroring production where even partially voiced bilabial stops were rarely produced with long post-release aspiration durations. Overall, the presence of aspiration affects the identification of voiced stops less in more posterior places of articulation, coinciding with the longer aspiration durations observed in production for velar stops compared to bilabial stops. Although these tendencies are seen in many participants, there remains a considerable amount of interindividual variability. Some listeners, for example, seem to require at least 40% of voicing in bilabial stops to choose the VOICED response, even when aspiration duration is short or zero. Similarly, some individuals do not consistently choose the VOICED response for the velar stop even when the closure is entirely voiced, if any aspiration is present. Such results underscore the diversity of participant responses on such a 2AFC perception task.

Intraindividual variability, measured here as response consistency, was highly similar across the four minimal pairs as seen in Figure 3.9 which presents the distribution of the Coefficient of Unalikeability (CoU) values for each of the four continua. The degree of variability was significantly correlated between minimal pairs with the same place of articulation (e.g. /g-kɛ/ - /g-ko/), across most other pairs (e.g. /g-kɛ/ and /b-pɛ/) and between aggregated values of the velar and bilabial continua, as reported in Table 3.4. Figure 3.10 shows the most significant correlation between the CoU of Velar and Labial categories. This suggests that participants' response consistency is a stable characteristic for the same acoustic cues, regardless of specific segment identity. This aligns with the production results where intraindividual variability was only correlated across segments utilizing the same acoustic ranges and articulatory configurations.

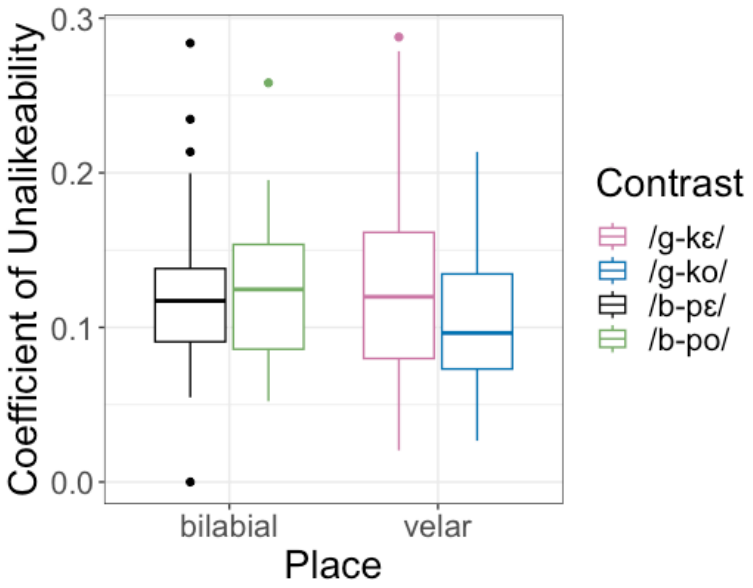


Figure 3.9 Distribution of values of Coefficient of Unalikeability (CoU) across speakers, indicating individual response consistency across different trials for the four stop voicing continua in the perception experiment. A low CoU indicates high consistency.

Table 3.4 Correlation coefficients and p-values between Coefficient of Unalikeability (CoU) values of all combinations of stop voicing continuum categories.

Contrast	**r**	**p**
CoU of **/g-kɛ/** – CoU of **/g-ko/**	.34	.02
CoU of **/b-pɛ/** – CoU of **/b-po/**	.47	< .001
CoU of **/g-kɛ/** – CoU of **/b-pɛ/**	.50	< .001
CoU of **/g-ko/** – CoU of **/b-po/**	.27	.06
CoU of **/b-pɛ/** – CoU of **/g-ko/**	.37	.009
CoU of **/g-kɛ/** – CoU of **/b-po/**	.52	< .001
CoU of **Velar Stops** – CoU of **Labial Stops**	.61	< .001

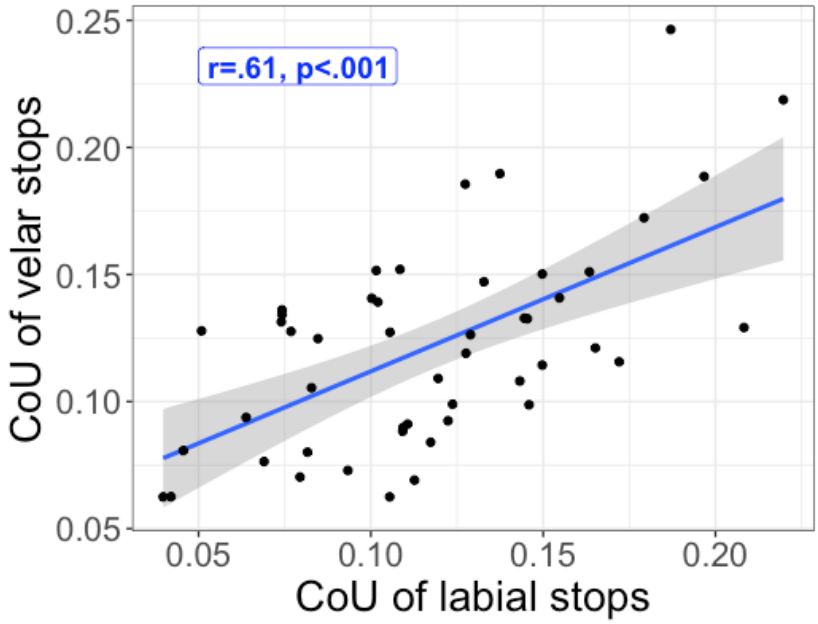


Figure 3.10 Significant Pearson correlation between mean the Coefficient of Unalikeability (CoU) of velar stops and the mean CoU of bilabial stops in the perception experiment.

3.2.2 Vowels

Aggregated results across all speakers are shown in Figure 3.11. Considerable interindividual variability was observed across both continua but was greater for the /u-o/ continuum. Specifically, a few listeners exclusively chose /o/ on the /u-o/ continuum and several listeners chose /ɛ/ for nearly all trials on the /e-ɛ/ continuum. Interestingly, these listeners who chose only one of the two options for all or nearly all trials are not the same across continua. In other words, a participant's behavior in the identification of one continuum does not necessarily predict their behavior on another continuum, even across similar phonological categories, such as vowel contrasts. The amount of intraindividual variability (response inconsistency) was similar across both continua as seen in Figure 3.12, however no correlation was observed between CoU on the /e-ɛ/ and the /u-o/ continuum across speakers ($p = 0.2$).

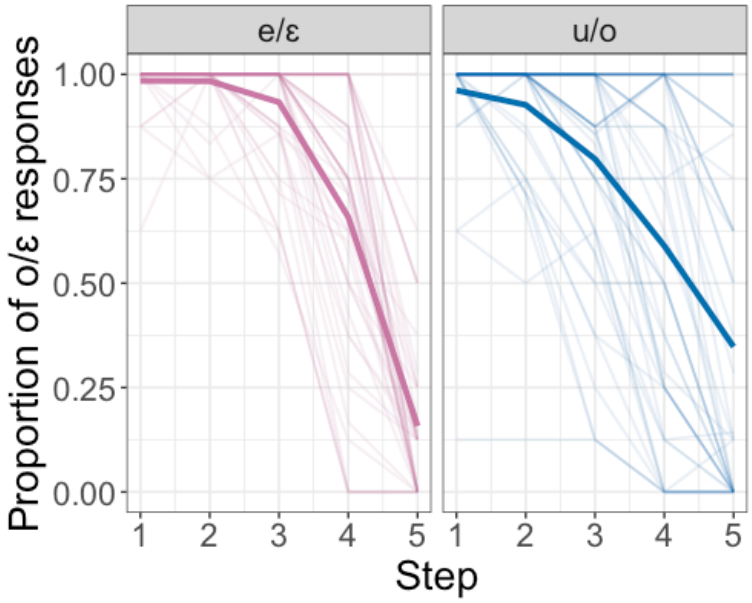


Figure 3.11 Proportion of /ɛ/ or /o/ responses for the two vowel continua in the perception experiment. Thick, dark lines represent the mean across all participants, while pale lines show individual trajectories.

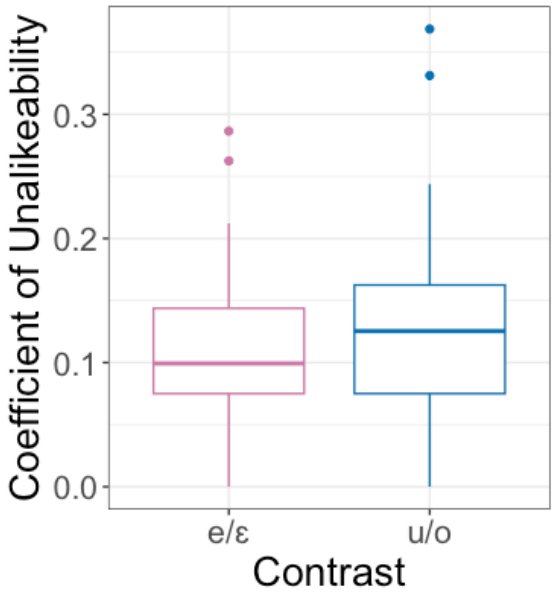


Figure 3.12 Distribution of values of Coefficient of Unalikeability (CoU), indicating individual response consistency across trials for the two vowel continua in the perception experiment. A low CoU indicates high consistency.

3.2.3 Fricatives

Overall, participants' responses to the fricative continuum could be grouped into three categories: those that chose /ʃ/ 95% of the time (n=29, henceforth *Only /ʃ/*), those that chose /s/ 95% of the time (n=5, henceforth *Only /s/*)and those whose responses followed the purportedly expected identification curve, choosing both /s/ and /ʃ/ (n=15, henceforth *Both /s-ʃ/*). Aggregated results across speakers in each of these three groups are shown in Figure 3.13. CoU values, representative of the degree of intraindividual variability, are presented in Figure 3.14. The high prevalence of a CoU value of 0 reflects the lack of variability observed in Only /ʃ/ and Only /s/ participants.

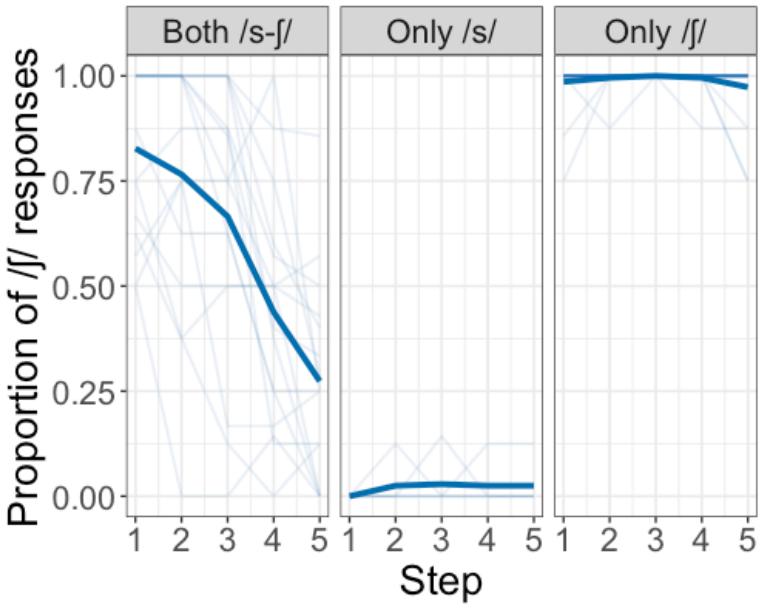


Figure 3.13 Proportion of /ʃ/ responses for the fricative continuum in the perception experiment. Thick, dark lines represent the mean across all participants, while pale lines show individual trajectories. Participants were divided into three categories based on their response patterns.

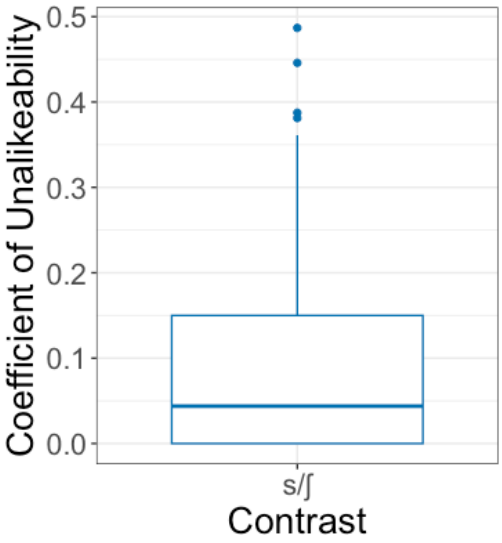


Figure 3.14 Distribution of values of Coefficient of Unalikeability (CoU), indicating individual response consistency across different trials for the fricative continuum in the perception experiment. A low CoU indicates high consistency.

It is intriguing that three such distinct behaviors were observed for the fricative continuum. Although considerable interindividual variability was present on the stop voicing and vowel continua, most speakers' responses aligned with a single group-level tendency. The division of the responses on the fricative continuum seem to indicate distinct approaches to the task, possibly due to participants paying attention to different aspects of the acoustic signal or interpreting the fricative continuum as a distinct task compared to the other continua.

It is possible that the use of synthetic stimuli affected listeners' responses. Makashay (2003) observed that phoneme identification, specifically the contrast between /θ-s/ and /i-ɪ/, is less accurate for non-speech sounds than natural speech sounds and that this difference is greater for fricatives than for vowels. Specifically, the majority of listeners in Makashay (2003)'s study never identified the non-speech fricative stimuli as /s/. In addition, a large degree of interindividual variability was observed, where some listeners categorized the non-speech stimuli similarly to natural stimuli, while others presented two distinct behaviors when responding to non-speech and natural speech stimuli.

Another possible factor that influenced participants' responses is the lexical frequency of the word choices presented to listeners. In the corpus of spoken French in Quebec (Dostie 2016; CFPQ 2006) which contains 712,300 words, the verb *lâcher*, including all possible conjugations, appears 60 times whereas *lasser* is entirely absent. It is therefore possible that the low lexical frequency of the word *lasser* in Laurentian French biased participants to choose the much more common word *lâcher*. Previous studies have shown that when presented with acoustically ambiguous segments, listeners more often report perceiving words with higher lexical frequency (Cooke et al. 2019). Although the possibility of such effects was considered when selecting the stimuli for this continuum, the other requirements of experimental design, namely that words should be existing Laurentian French words and should constitute a true minimal pair, were highly limited by the phonotactics of the Laurentian French lexicon.

In addition to lexical frequency effects that can bias listeners' perception, contextual information can also have an important effect. For example, Jongman et al. (2003) observed that identification of natural fricatives in English was greatly affected by semantic cues. Specifically, listeners correctly identified the fricative with greater accuracy when the sentence in which it was embedded was semantically congruous, i.e. where only one of the provided fricative options produced a logical sentence. Shastri et al. (2019) argued that a lack of context renders perception difficult for some individuals. In their study, Malayalam listeners' accuracy rate in the perception of natural native phonemes embedded in nonce words varied widely, ranging from 58%-100%. Interestingly, accuracy rate was highly correlated with various measures of working memory capacity, indicating that phoneme identification without contextual cues can increase cognitive load and may be a particularly difficult task for some participants.

Most likely, the results for the fricative continuum are due to a combination of the previously mentioned factors. Only /ʃ/ listeners may have paid more attention to frequency effects than spectral cues, possibly due to the synthetic nature of the fricative stimuli, causing them to employ a different strategy in the identification task. Only /s/ listeners on the other hand may have also found the task difficult due to the synthetic nature of the stimuli, but possibly gave preference to phoneme, rather than lexical, frequency. Indeed, /s/ occurs with a much higher frequency in French than /ʃ/. Across all positions, in the LEXIQUE database (New et al. 2004), /s/

appears 17,915 times compared to only 2,732 for /ʃ/ (Piot et al. 2024). It seems plausible that even if speakers were presented with a relatively unambiguous token, the very low lexical frequency of *lasser* or the low frequency of the phoneme /ʃ/ itself, caused them to ignore the mismatching information and base their response on non-acoustic information.

The Both /s-ʃ/ listeners, on the other hand, may have been able to ignore lexical effects and tacitly accept the synthetic nature of the stimuli in order to base their responses on spectral cues. Overall, the large amount of variability, and more importantly the distinctly opposite behaviors or certain groups of participants, is an important result and highlights the pertinence of considering individual tendencies in interpreting experimental data. Such results of different underlying participant behaviors have been observed in various perception studies. Lai et al. (2022), for example, observed a trade-off relationship across participants between the influence of coarticulatory information and lexical cues in a fricative /s-ʃ/ continuum task. They suggest this reflects individual differences, where some listeners naturally rely more heavily on lexical information than coarticulatory cues and vice versa. Giovannone & Theodore (2021) observed that individuals with weaker language ability, as measured by standardized tests evaluating expressive and receptive language skills, relied more heavily on lexical information than acoustic cues. Even more relevant to the results of this dissertation is the observation by Zhai et al. (2026) who show that the tendency to rely on lexical or acoustic information appears to be specific to the continuum tested and does not represent an overall speaker trait.

3.2.4 Variability across continua

Figure 3.15 presents the CoU values for each of the four continua on which participants were tested in the perception experiment: stop voicing, /e-ɛ/, /u-o/ and /s-ʃ/. Given that intraindividual variability was highly correlated across all four stop voicing continua (/b-pɛ/, /b-po/, /g-kɛ/, /g-ko/), a single mean CoU value was calculated for the stop voicing category. For most participants, the lowest degree of intraindividual variability is observed for the fricative contrast, largely due to the majority only making use of one of the two choices. The variability observed across the other continua is similar in magnitude. A linear regression with CoU as the

independent variable, the acoustic contrast as the predictor (Velar stops, Labial stops, /e-ɛ/, /u-o/ and Fricative) and a random by-subject intercept revealed no significant effects (p>0.1), indicating that, at the group level, the degree of intraindividual variability does not depend on the type of acoustic contrast (Table A22).

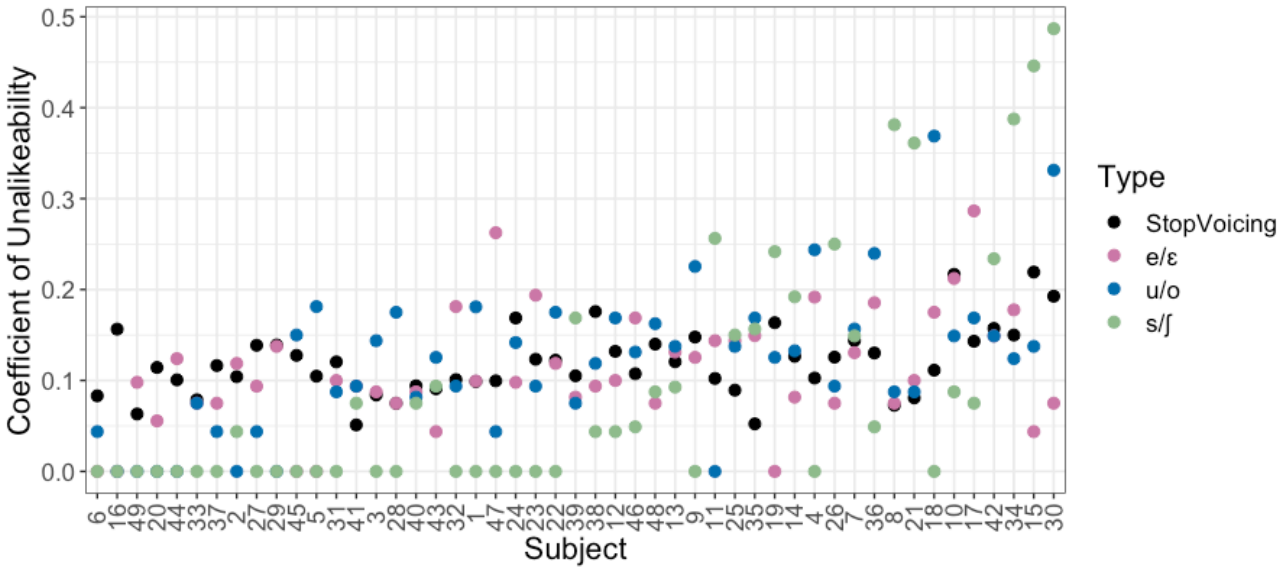


Figure 3.15 Coefficients of Unalikeability (CoU) for each of the perception continua tested, arranged by participants with lowest (leftmost) to highest (rightmost) overall variability.

Additionally, Pearson correlations carried out between the mean Coefficients of Unalikeability across phonological contrasts (Vowels-Stops, Vowels-Fricative, Stops-Fricative) were not significant (p>0.1), indicating that at the individual level, the degree of inconsistency in the perception of one phonological contrast does not predict the degree of inconsistency in other contrasts. Considering that the acoustic measures tested here were of a different nature (largely durational for stop voicing, low-frequency spectral for vowels and high-frequency spectral for fricatives), this result coincides with the finding that production variability is dependent on the specific acoustic cue being tested and does not represent an across-the-board individual characteristic. This also corresponds with previous studies where no correlation in response consistency was observed across distinct acoustic continua (Fuhrmeister & Myers 2021; Zaharchuk et al. 2026), but clearly differs from studies that did observe a significant correlation

(Myers et al. 2024; Kim et al. 2025). It is at this point unclear what the cause for such conflicting results may be, however it is possible that additional factors, such as language background or demographic characteristics, may determine the presence or absence of cross-segment correlations in response consistency. It is also just as likely that individuals employ different strategies across perception tasks and it is thus difficult to compare results across experiments that analyzed different acoustic contrasts using distinct methodologies, such as the 2AFC in the current dissertation and the VAS task in the previously mentioned studies.

Given these conflicting results from previous studies and the large amount of interindividual variability observed in the perception experiment, in particular for the fricative continuum, it is pertinent to consider how additional factors, such as sociodemographic or language-background characteristics, may have affected participants' performance. Although sociodemographic factors were not directly controlled for during participant recruitment, participants naturally differ from one another across place of origin, gender, and linguistic experience. As explained in Section 2.1.1, all participants in the current study spoke at least one additional language alongside their native language, Laurentian French, and originated from and spent most of their lives in either Quebec or Ontario. Participants differed little from one another in age but varied considerably in the age of acquisition and proficiency of their L2 and the presence of an L3. All of the factors considered are summarized in Table 3.5. It should be noted that L2 (for the variable *Age of acquisition of L2*) refers to English in non-native bilinguals and to a third language in native bilinguals, who are considered to have two L1s. Native bilinguals who had not learned a third language were excluded from the *Age of acquisition of L2* variable. The role of such factors in predicting the results presented in Chapter 2 for Production B will also be explored here.

Independent Welch t-tests were used to quantify the effect of the categorical variables, whereas Pearson correlation analyses were used to investigate the role of the numerical variables, as specified in Table 3.5. Analyses of possible interactions were not considered as this led to highly uneven distributions across groups. Dependent variables from the perception experiment were the Coefficient of Unalikeability (CoU) measures representing response variability in the perception task for the four continua tested: stop voicing, /e-ɛ/, /u-o/ and /s-ʃ/.

For the production experiment, the dependent variables were the Coefficient of Variability (CoV) measures for stop voicing (voicing proportion and post-release aspiration duration, both averaged across place), vowel formants (F1 and F2 for each of the four vowels) and fricative COG (separately for /s/ and /ʃ/). Given the nature of sociodemographic and linguistic variables, many of these variables are likely not entirely independent of one another. Unfortunately, the size of the participant sample and its unbalanced nature did not allow a decorrelation of these sociodemographic and linguistic factors through a principal component analysis.

Table 3.5 Levels and counts of sociodemographic and linguistic variables. Categorical variables are written in italics and values in parentheses represent subject counts. Numerical variables are marked in bold with ranges are given in parentheses.

Variable	Levels
Native bilingual	*Yes (16) – No (33)*
Province of origin	*Ontario (23) – Quebec (26)*
English proficiency	**5 values (native – low)**
Age of acquisition of L2 (n=38)	**14 values (3 – 18)**
Age	**11 values (18 – 45)**
Gender	*Female (35) – Male (13) – Non-binary (1)*
Presence of L3	*Yes (20) – No (29)*
Proficiency of L3 (n=20)	**4 values (high – low)**
Time spent outside Ontario/Quebec	*0 years (45) – 1-2 years (2) – 3-4 years (2)*
Percentage of daily use of Laurentian French	*80-100% (24) – 50-80% (20) – 20-50% (5)*

Overall, most sociodemographic and linguistic factors could not predict or explain the patterns of variability observed in the production and perception experiments (p's > .2). It was found, however, that native bilinguals exhibited higher rates of variability than participants whose sole native language was Laurentian French on both vowel continua in the perception experiment (**/e-ɛ/:** β = 0.04, t = 2.3, p = .03, **/u-o/:** β = 0.07, t = 3.3, p = .001) and higher

intraindividual variability of F1 of /o/ (β = 0.02, t = 2.4, p = .02), F2 of /e/ (β = 0.01, t = 2.3, p = .03) and F2 of /ɛ/ (β = 0.01, t = 3, p = .004) in production. This is in line with certain studies that have observed less compactness in the production and perception of vowels in the speech of early bilinguals (Lara & Pillot-Loiseau 2023; Kogan 2024; Li et al. 2026), but contradicts Georgiou & Giannakou (2024), who reported that bilinguals have tighter vowel distributions than monolinguals. The exact relationship between variability and native bilingual status has been shown to depend on various factors not controlled for in this study, such as frequency of language use (Mora et al. 2015), exposure to different dialects (Levy & Hanulíková 2019) and code-switching habits (Li et al. 2026). Although this is an interesting result, it lies outside the scope of the current study, and given the relatively small number of native bilinguals, it should be interpreted with caution.

An additional significant correlation was observed between variability in the perception of the /u-o/ and the fricative continua, as measured by CoU and Age of acquisition of L2, as plotted in Figure 3.16. A near-significant, weak correlation was also observed for the stop voicing continuum, but no similar correlation was found for the /e-ɛ/ continuum.

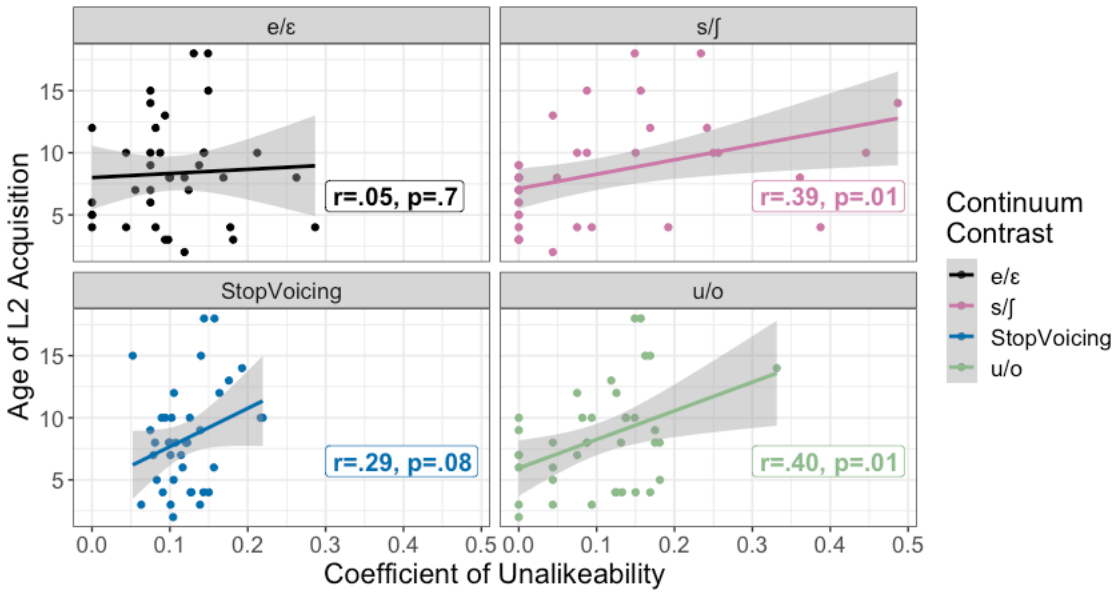


Figure 3.16 Pearson correlations between Coefficients of Unalikeability (CoU) of the four perception continua and the age of L2 acquisition.

It is quite interesting that the significant correlation between Age of L2 acquisition and intraindividual variability in the perception experiment was observed specifically for the fricative and the u/o continuum, as it was only on these two continua that some participants consistently chose the same response across all repetitions and stimuli which may indicate greater task difficulty. Although the current data only allow speculation, it appears that having learned a foreign language at a later age has the potential to raise a participant's attention to acoustic detail, possibly since they would have frequently been exposed to unfamiliar sounds in the L2 and would have had to develop more conscious strategies to effectively categorize these unfamiliar sounds. A similar result was observed by Blanco et al. (2016) who measured categorization accuracy of nonce words containing phonemes unique to either English or Spanish in monolinguals, early bilinguals and late bilinguals. Their results showed that late bilinguals achieved greater accuracy than both early bilinguals and monolinguals, suggesting a possible greater awareness of acoustic differences. It is possible that participants who began learning a foreign language later in life are more familiar with the task of distinguishing between two sounds, possibly due to their experience of carrying out similar relatively unnatural tasks in language classes. Additionally, in real-world situations, late learners of foreign languages likely have to pay more attention to acoustic details to differentiate sounds not found in their native languages, offering them an experience-based advantage in a task such as the 2AFC identification task used here. It is important to note, however, that age of L2 acquisition is only one of many aspects that have been shown to influence an individual's production and perception of speech sounds, be it in their native or a foreign language. The current amount of L2 or L1 exposure, the manner of L2 acquisition, L2 proficiency and many other factors would need to be taken into account to fully understand if there is truly a causal relationship between the age of L2 acquisition and the rate of intraindividual variability in the perception task.

An additional factor that can provide greater insight into participants' approaches and decision processes is reaction time. Mean reaction times for each of the four continua (the four stop contrasts were again grouped into one stop voicing continuum) were calculated for each participant. Pearson Correlations revealed that greater reaction times were associated with significantly greater CoU values for the fricative continuum, as shown in Figure 3.17. A similar

tendency was observed for both vowel continua, though the correlation failed to reach significance. No correlation was found between CoU of the stop voicing continuum and reaction time. The complete lack of correlation between variability and reaction time for the stop voicing continuum is likely due the difference in task type when compared to the vowel and fricative continua. The stop voicing continua manipulated two acoustic cues and were aimed at testing cue weights, whereas the vowel and fricative continua tested categorical boundary location.

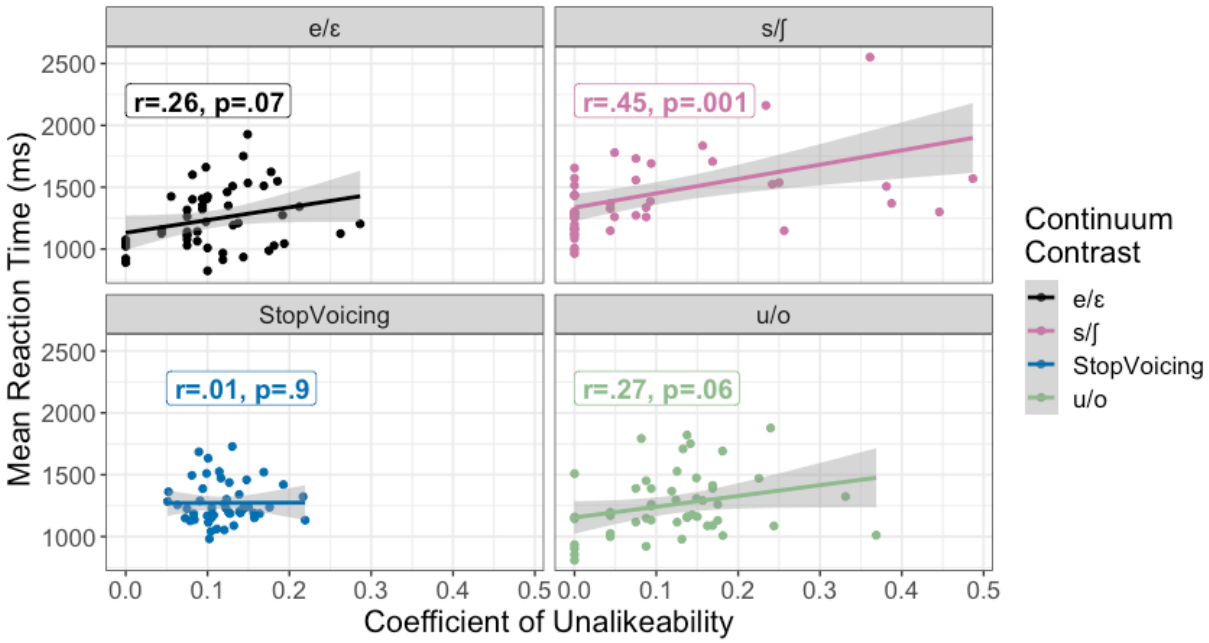


Figure 3.17 Pearson correlations between mean reaction times (ms) of the four perception continua and the corresponding Coefficient of Unalikeability.

Longer reaction times can be taken as indirect evidence of two very different types of reactions to a task: heightened attentiveness or distraction (Read et al. 2022). These distinctly opposite behaviors cannot necessarily be teased apart solely with data from the 2AFC experiment, although analyzing the production-perception link may offer a way to better understand individual participant strategies, as will be explained in Section 4.5. However, it is immediately apparent in Figure 3.17 that the majority of participants with CoUs of 0 had, on average, much lower reaction times. A CoU of 0 is representative of total response consistency across all eight repetitions of all stimuli for that continuum. For the fricative continuum, this was observed

almost exclusively in listeners who always chose the same binary response. It is also well-known that on difficult tasks, lower reaction times are indicative of decreased accuracy (Hui & Jia 2024). Although for a perceptual identification task as the one carried out in this dissertation, there is no true accuracy measure, a total lack of use of one of the two categories can possibly be seen as poorer engagement with the task. Therefore, it seems plausible that the Only /ʃ/ and Only /s/ listeners (whose mean reaction times were: Majority /ʃ/ = 1292 ms, Majority /s/ = 1384 ms) made less of an effort to accurately categorize the stimuli they heard on the fricative continuum and their short reaction times are an indicator of weakened task-attentiveness. Listeners in the Both /s-ʃ/ group had much longer mean reaction times (mean = 1762 ms) and greater values of CoU, likely indicating they paid more attention to the task. Differing degrees of task-attentiveness will be further explored after examining the link between production and perception in Section 4.5. Again, the influence of external factors such as lexical frequency and stimuli naturalness, discussed in Section 3.2.3, likely also played a role in raising the degree of ambiguity of the fricative stimuli. When faced with this greater ambiguity, some listeners likely decided to expend less effort and thus responded less attentively.

Finally, it is important to note that research has repeatedly shown that phonetic perception is a skill like any other, meaning some people are naturally more sensitive than others and achieve higher levels of accuracy in distinguishing linguistic sounds (Díaz et al. 2008; 2016). Given the large amount of interindividual variability observed in the current study, both in the categorization of the two phonemic categories and in the rate of response consistency across continua, it is clear that speech perception is a highly individual characteristic of language users and it is probable that listeners develop unique, individual strategies to achieve accurate perception in daily language use. Additionally, the large degree of interindividual and intraindividual variability in task approach, also observed by Zhai et al. (2026), are indicative that the type of task itself can greatly affect the results obtained, suggesting that the results of a 2AFC perception task may represent different aspects of perception for different individuals. The implications of such a possibility will be further explored in Section 5.2.

3.3 Conclusion

Considerable intraindividual variability, corresponding to response consistency across repetitions, was observed for all continua. Coinciding with the results for production variability, the degree of variability in perception was unique to the acoustic contrast tested, confirming that intraindividual variability is not a general characteristic of an individual but is specific to certain phonological categories and acoustic properties. This suggests that speech perception cannot be seen as a global skill but rather that each perception continuum is potentially treated as a distinct task. Large interindividual differences were also observed for all continua, indicating that individuals differ widely from one another in their weighting of voicing proportion and post-release aspiration duration on the stop voicing continuum, in their categorization of phonological segments on the vowel and fricative continua, and in the degree of response consistency for any given continuum. These results underline the importance of assessing individual differences, as group tendencies can potentially hide distinct participant strategies. The relationship between acoustic cue weights, phonological categories and degree of intraindividual variability across production and perception will be explored in Chapter 4.

4. Link between production and perception

In addition to analyzing individual patterns in production and perception, it is important to understand the relationship between these two speech modalities and how acoustic values and intraindividual variability observed in one dimension may be able to predict or inform observable patterns in the other dimension. As described in Section 3.1, two types of continua were employed in the perception experiment. The first type was the stop voicing continua (covering four similar contrasts: /b-pɛ/, /b-po/, /g-kɛ/, /g-ko/) where two covarying acoustic properties, voicing proportion and duration of post-release aspiration, were manipulated. The second type was the continua for vowels (across two contrasts: /e-ɛ/ and /u-o/) and fricatives (/s-ʃ/,) where only one aspect of the acoustic signal was modified (F1 and F2 for vowels and COG for fricatives). Due to this, the perception results for the two types of continua differ in how they relate to phonological representations: the stop voicing continua tested the relative weight of the two voicing cues, while the vowel and fricative continua directly tested categorical boundaries between two phonological categories. Given these inherent differences, two distinct methods were employed to operationalize the link between production and perception, as will be explained in the following sections.

4.1 Voiced and voiceless stops

4.1.1 Methodology

As described in the previous chapters, the analysis of voiced and voiceless stops in both production and perception investigated the use of two cues: post-release aspiration duration and voicing proportion, observed as voicing wane (voicing dying out in the *final* third of the closure) in voiced stops and carryover voicing (voicing dying out in the *first* third of the closure) in voiceless stops. To understand how production and perception may be related, the relative importance of each of these cues in both modalities needs to be established.

Following previous research on interindividual variability in perception (Kong & Edwards 2016; Clayards 2018; Honda et al. 2024a, among others), cue weights for voicing proportion and

post-release aspiration duration were determined by extracting by-subject random slope coefficients from the mixed effects logistical regression model presented in Section 3.2.1 and Table A21. These individual beta (β) coefficients represent the increase in log-odds of a VOICED response in the 2AFC perception task as the acoustic variable (voicing proportion or post-release aspiration duration) increases. Beta coefficients can only be used as reliable indicators of individual cue weighting strategies if the logistic regression residuals are normally distributed and the model predictions are linear. Both of these conditions were met by the model presented in Section 3.2.1.

To determine a similar coefficient for production values, Cohen's ds (Cohen 1988) were calculated, as is common in research on production variability (Tạ et al. 2022; Brunelle 2022, 2024, among others). Cohen's ds represent the degree of separation between two categories (here voiced and voiceless stops) and were calculated separately for each subject by means of the following formula:

$$d = \frac{M_a - M_b}{SD_{(a+b)}}$$

In this formula, *M* refers to the mean, *a* and *b* refer to the groups containing voiced and voiceless stops, respectively, and *SD* is the standard deviation of voiced and voiceless stops pooled together. A Cohen's d greater than 0.8 is generally taken as corresponding to a large effect size, indicating a great degree of separation between the two categories. Individual Cohen's ds were calculated for voicing proportion and post-release aspiration duration separately, using values from Production B.

4.1.2 Results

Considerable interindividual variability was observed in both beta coefficients, as shown in Figure 4.1, and Cohen's ds, as shown in Figure 4.2. In Figure 4.1, a greater absolute value of a beta coefficient indicates that an increase of the given acoustic property had a greater influence

on the likelihood of participants' selection of the VOICED response. Overall, an individual's perception cue weight for voicing proportion is of a similar magnitude as their cue weight for aspiration duration, emphasizing that at the individual level, similar sensitivy is observed for both acoustic properties.

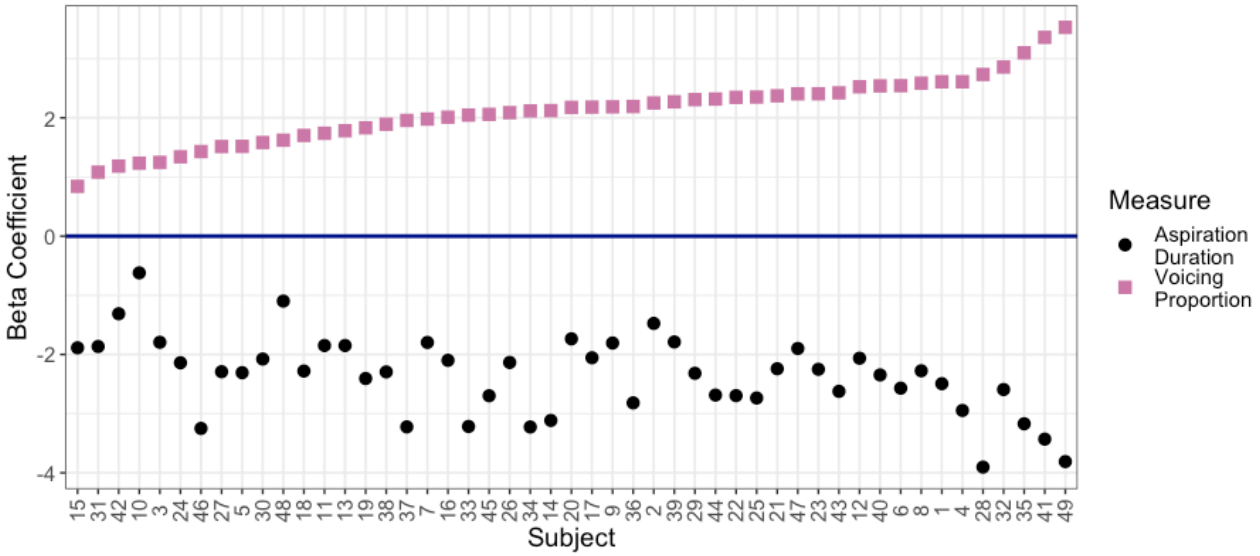


Figure 4.1 Individual by-subject random slope (beta) coefficients extracted from a mixed-effect logistic regression model, representing the likelihood of choosing the VOICED category as either voicing proportion or aspiration duration increased. Subjects are arranged from lowest (leftmost) to greatest (rightmost) beta coefficient of voicing proportion.

In Figure 4.2, a positive Cohen's d indicates an association with the phonological category of voiced stops and a negative Cohen's d indicates an association with the phonological category of voiceless stops. It is unsurprising that Cohen's ds for voicing proportion are quite large, given the large differences between average values of voicing proportion in voiced and voiceless stops. While for some participants, aspiration duration significantly divided the categories of voiced and voiceless stops, the effect size was much smaller than for voicing proportion across the board. This corresponds to the observation that many speakers consistently produced aspiration during the realization of voiced stops, implying that aspiration duration alone is not as strong a cue as voicing proportion. Overall, voicing proportion is the dominant cue used in the distinction of voiced and voiceless stops for most individuals.

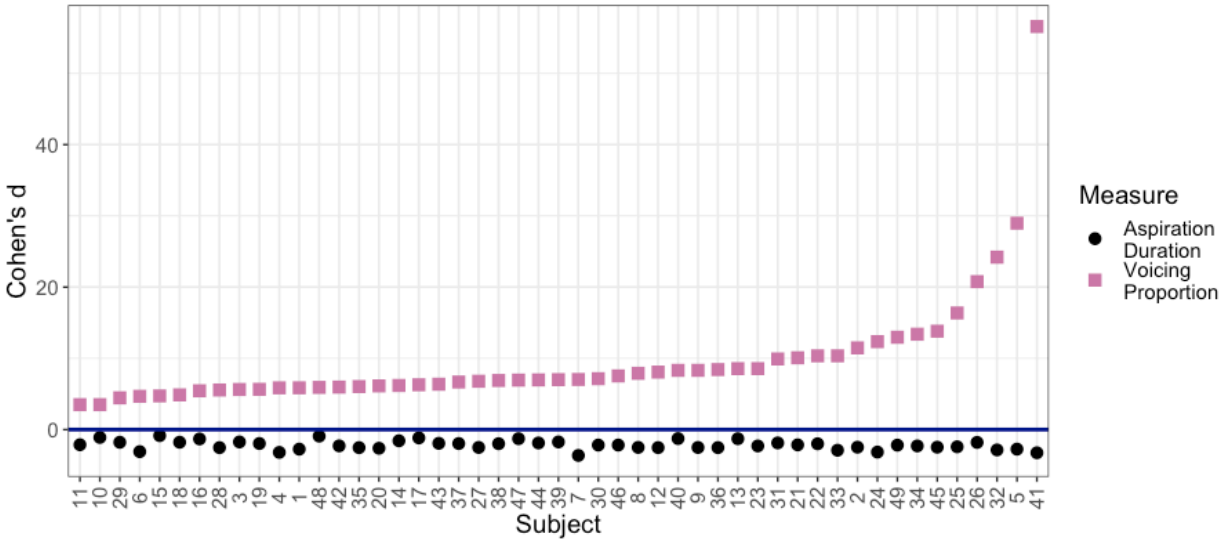


Figure 4.2 Individual Cohen's ds by subject. A positive value indicates that higher values of that measure are associated with phonologically voiced stops and a negative value indicates that higher values of that measure are associated with phonologically voiceless stops. Subjects are arranged from lowest (leftmost) to greatest (rightmost) Cohen's d of voicing proportion.

A few speakers had extremely high Cohen's ds, due to very low standard deviations across both categories of stops (nearly all phonologically voiced stops were fully voiced and nearly all phonologically voiceless stops were completely voiceless). For the sole purpose of facilitating visualization of the relative importance of cues and interindividual differences, Cohen's ds were log-transformed. Log-transformation essentially compresses the range of large values and expands the range of small values to approach a more normal distribution. Negative Cohen's ds were multiplied by -1 before log-transformation and again by -1 after to maintain the correct graphical representation. The log-transformed values are shown in Figure 4.3.

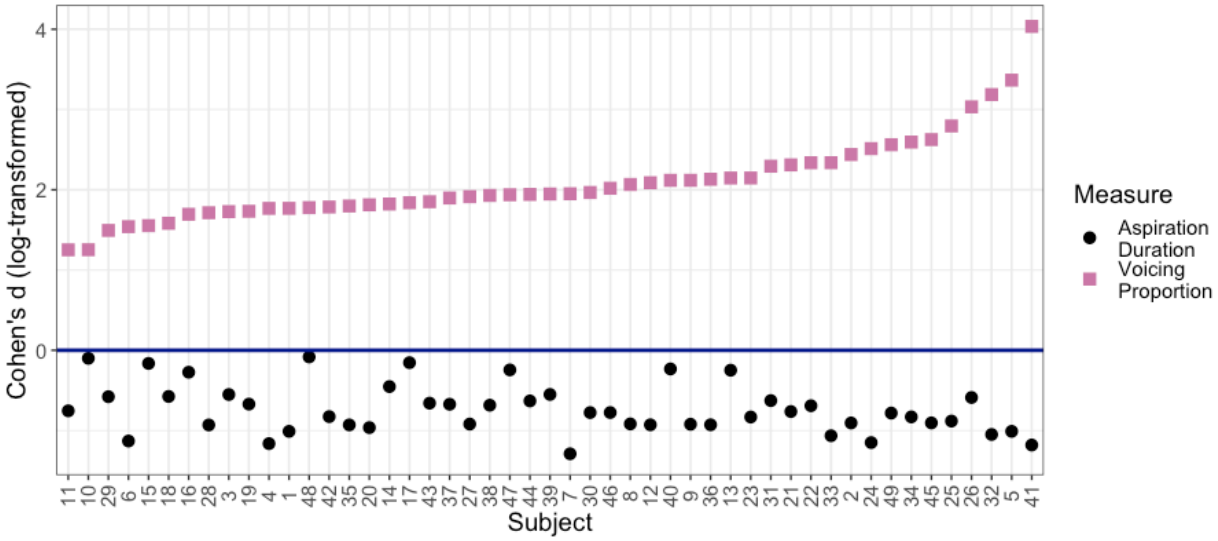


Figure 4.3 Individual log-transformed Cohen's ds by subject. A positive value indicates that higher values of that measure are associated with phonologically voiced stops and a negative value indicates that higher values of that measure are associated with phonologically voiceless stops. Subjects are arranged from lowest (leftmost) to greatest (rightmost) Cohen's d of voicing proportion.

In order to directly compare production and perception weights, all Cohen's ds and beta coefficients were normalized by scaling values to be contained between 0 and 1 using the following formula, where X is the absolute value of the Cohen's d or beta coefficient value:

$$X_i = \frac{X - X_{min}}{X_{max} - X_{min}}$$

The original Cohen's d values were used here and *not* the log-transformed values shown in Figure 4.3. The absolute values of beta coefficients and Cohen's ds were used because the mathematical sign of both measures is solely dependent on the arbitrary decision of which category is assigned to positive values (in this study it is the phonological category of voiced stops). The effect size (for Cohen's ds) or log-odds (for beta coefficients) itself is sign-independent.

Using these normalized measures, Pearson correlations showed that, at the group level, the respective importance of both acoustic properties for stop voicing is significantly correlated

across modalities, as shown in Figure 4.4. This implies that participants with more distinct categories in production were more sensitive to these cues in perception. Individuals who produced less partial voicing thus accepted more instances of partially voiced tokens as phonologically voiced stops in perception. Similarly, speakers who produced less phonologically voiced stops with aspiration weighed aspiration more heavily in perception as being a characteristic of phonologically voiceless stops. As always, there remains a considerable degree of interindividual variability.

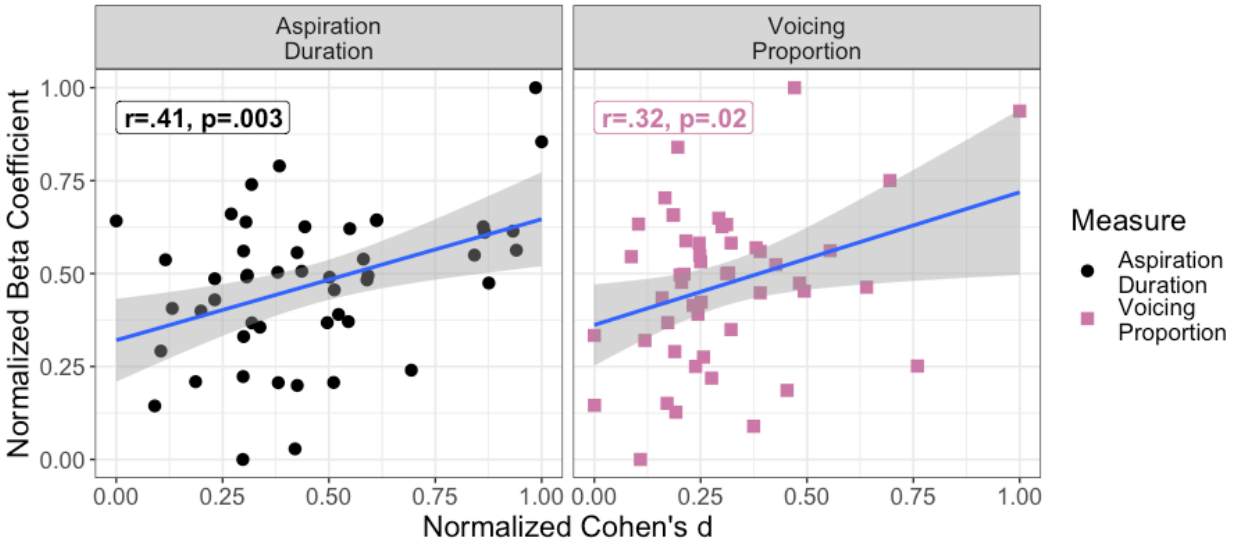
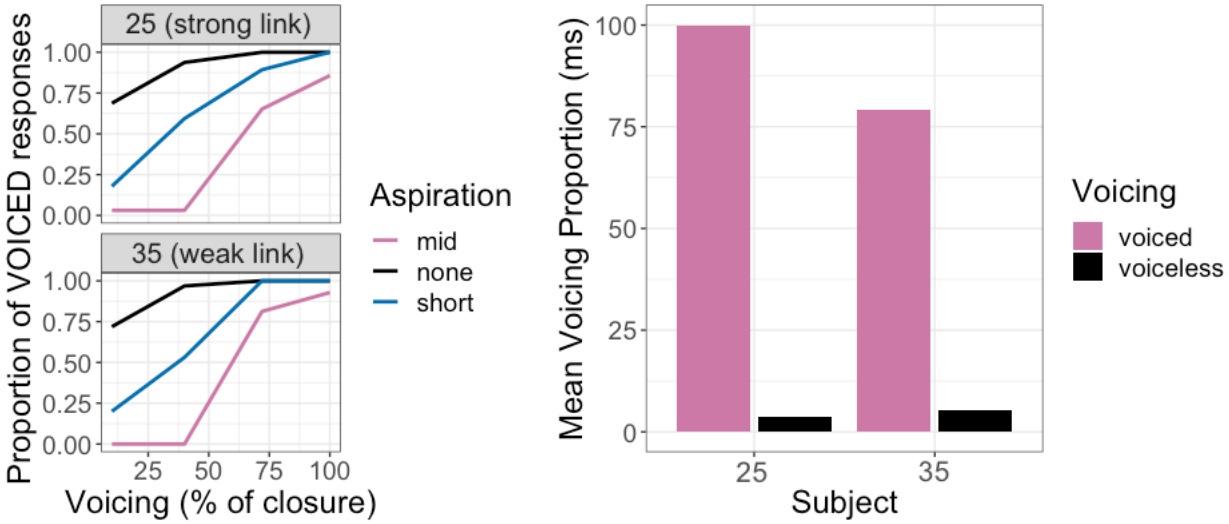


Figure 4.4 Significant correlations between normalized (scaled to a range of 0-1) Cohen's ds, representative of production weights, and beta coefficients, representative of perception weights.

To understand the relationship between modalities, a production-perception link coefficient was calculated for each speaker by taking the absolute value of the difference between the normalized Cohen's d and normalized beta coefficient for each of the two measures, similar to Noiray et al. (2014)'s measure of acoustic-articulatory similarity in English vowels. This measure is referred to as the *production-perception difference* or *production-perception link coefficient*. A smaller production-perception difference indicates greater similarity between the two modalities.

In order to illustrate the interpretation of the production-perception difference, examples of participants with strong and weak production-perception links are presented in Figure 4.5, for

voicing proportion and Figure 4.6, for aspiration duration. In Figure 4.5, participant 25 produced exclusively fully voiced *voiced* tokens and weighed the voicing proportion cue highly in perception, although conditioned by aspiration duration. This indicates that for this individual, the presence of voicing is strongly associated with voiced stops in both modalities. Participant 35 also weighed voicing proportion heavily in perception, with nearly identical responses as participant 25. However, this speaker produced nearly all voiced stops with partial voicing, reaching on average only 75% of the closure duration. Participant 35's production-perception difference is therefore much higher, indicating a mismatch between the importance of the voicing proportion cue in each modality.

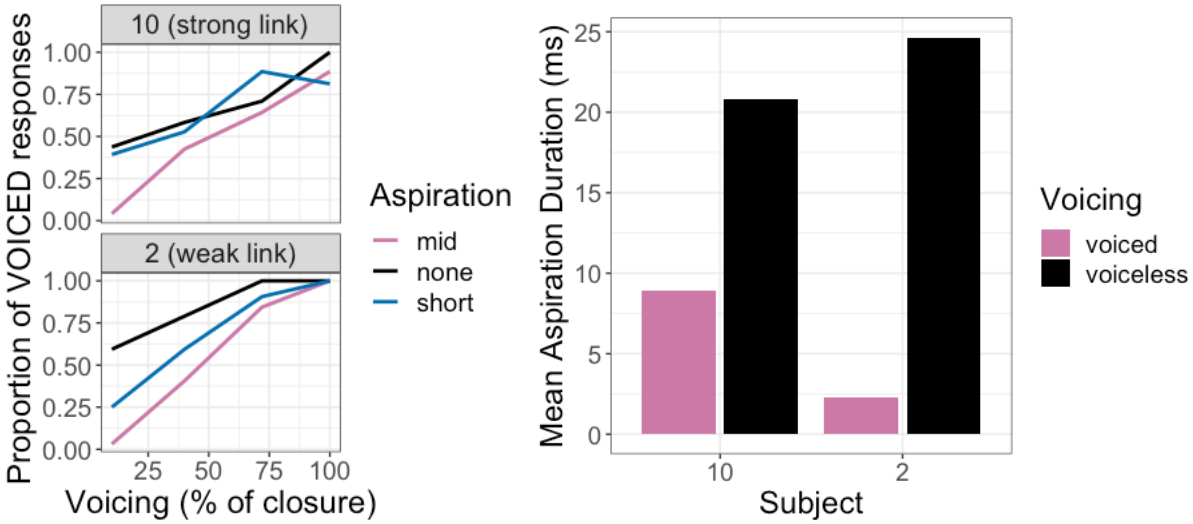


Production-perception difference for Voicing Proportion	
Participant 25 = 0.007	Participant 35 = 0.64

Figure 4.5 Two example participants, one with a strong production-perception link (25) and one with a weak production-perception link (35), for the acoustic property of voicing proportion.

In Figure 4.6, the probability of a VOICED response in participant 10's results are quite similar across the three levels of aspiration, indicating low importance for this cue in perception. Likewise, in production, participant 10 consistently produced both voiced and voiceless stops

with aspiration. Therefore, the strong production-perception link for participant 10 indicates that aspiration duration plays only a small role in the distinction of voiced and voiceless stops in both modalities. On the other hand, participant 2 rarely produced aspirated voiced stops and even when doing so, still maintained a large difference between the two categories. However, in perception, the probability of a VOICED response is affected only slightly by an increase in aspiration; indeed, all three aspiration steps follow the same line and are quite close to one another. The role of aspiration duration in production and perception thus differs considerably for participant 2.



Production-perception difference for Aspiration Duration	
Participant 10 = 0.01	Participant 2 = 0.42

Figure 4.6 Two example participants, one with a strong production-perception link (10) and one with a weak production-perception link (2), for the acoustic property of post-release aspiration duration.

The distribution of the production-perception difference for stop voicing across all individuals is shown in Figure 4.7, evidencing a large degree of variability in the strength of the production-perception link across participants and acoustic cues. Although for some participants, the strength of the production-perception link was of a similar magnitude for both voicing

proportion and aspiration duration, there was no group-level correlation between these two coefficients (p = .2). As discussed in Section 2.2.1.3 and Section 2.3, closure voicing and post-release aspiration correspond to two distinct articulatory configurations and acoustic consequences. The lack of correlation between production-perception differences of the two cues suggests that the strength of the production-perception link is not a general individual characteristic but rather dependent on the acoustic property considered. This is in line with the observation that both production and perception intraindividual variability are highly dependent on the phonological segment and acoustic cue considered.

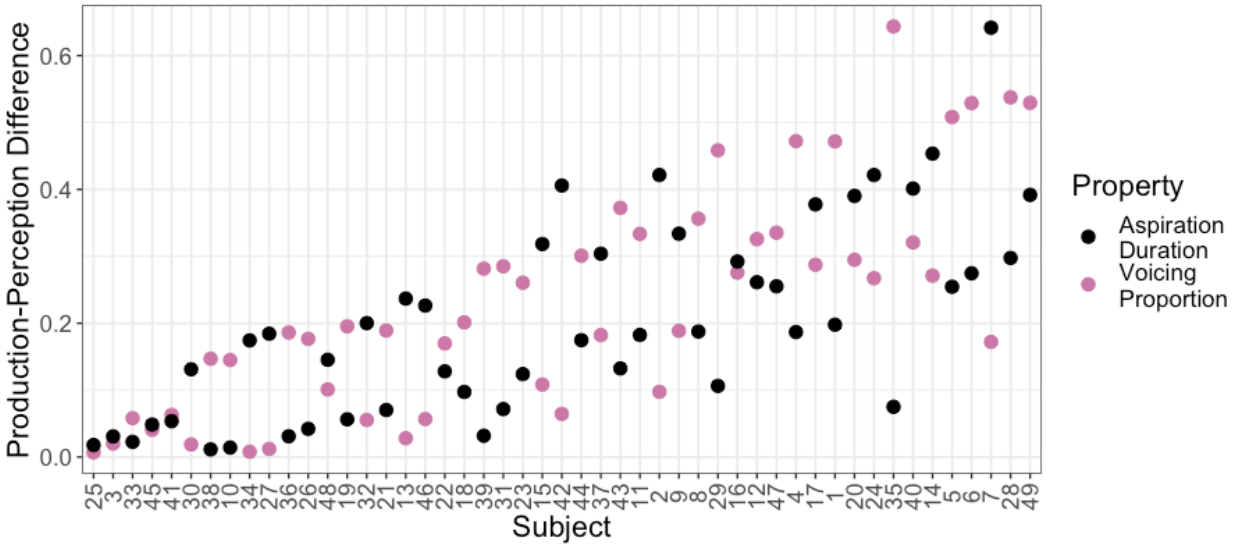


Figure 4.7 Production-perception link coefficients, arranged by degree of overall link across both acoustic properties, from those with the strongest (leftmost) to those with the weakest (rightmost).

Given the large amount of interindividual variability of the production-perception difference, it is important to consider how intraindividual variability may affect the strength of this link. A Pearson correlation revealed that greater response inconsistency on the stop continuum in the perception task (measured by the Coefficient of Unalikeability) was correlated with a stronger production-perception link for the voicing proportion cue but not for aspiration duration, as shown in Figure 4.8. This result seems to indicate that participants who are less consistent in the perception experiment make more active use of their own production cues

during the task. In other words, these individuals compare each stimulus in real time to the probability distribution of their own cue weights and choose what fits the best. They may carry out this process each time anew, thus leading to different responses across identical trials and greater overall variability. This possibility will be further explored in Section 4.5 and Section 4.6.

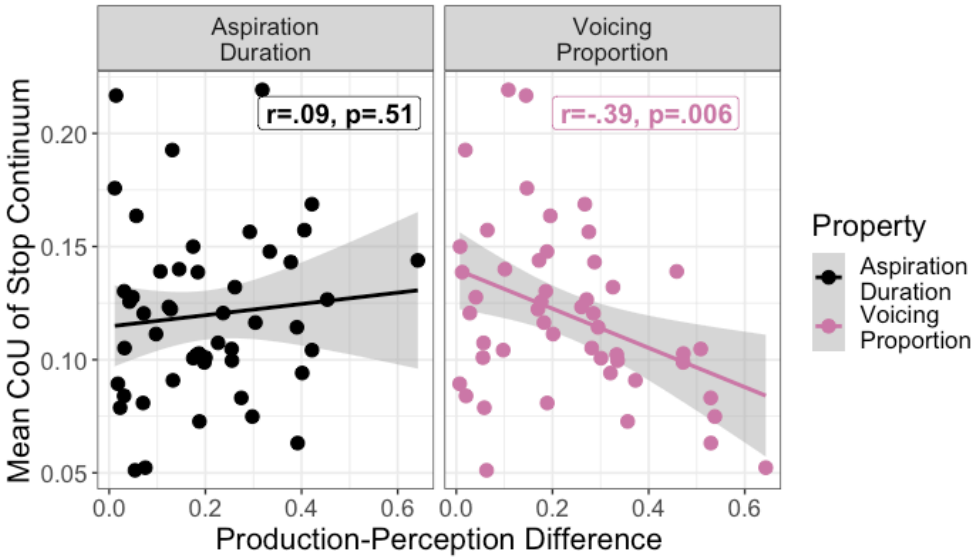


Figure 4.8 Pearson correlations between mean Coefficients of Unalikeability (CoU), measuring response consistency on the stop voicing continuum in the perception task, and the production-perception link coefficient. A lower value on the x-axis indicates a stronger production-perception link.

The presence of this correlation for voicing proportion but not for aspiration duration may be due to the overwhelming importance of the voicing proportion cue observed in production, as shown in Figure 4.3. Listeners who more actively used their production patterns when performing the 2AFC identification task may have thus prioritized the more salient of the two cues available to them when making their decision. It should be noted, however, that the individual cue weights extracted from the random slope coefficients may be influenced by response consistency, at least for some participants. In other words, more indecisive individuals, who also tend to have a higher CoU in the perception experiment, often have weaker cue weights, though this is not always the case. Given that the production-perception link coefficient is determined by values from both production and perception, the significant correlation in Figure

4.8 likely represents a real tendency, but it should still be interpreted with caution. It will be important to compare this result with those from the other phonological categories tested to better evaluate its significance.

No attempt was made to probe a comparable correlation between the CoV and the production-perception link. This was because both the production weights (Cohen's ds) used to calculate the production-perception link coefficient and the Coefficient of Variation (CoV) representing production variability were calculated using means and standard deviations of the same acoustic measures in production.

4.2 Vowels

4.2.1 Methodology

As the vowel continua varied across a single acoustic dimension (formant structure) participants' responses cannot be taken as an indication of cue weights. Therefore, the method used for the stop voicing contrast in analyzing the production-perception link is not applicable here. The vowel continua, as well as the fricative continuum, probed the location of participants' categorical boundaries in perception. The location of this categorical boundary was thus determined in production and perception for each individual and the production-perception link was operationalized as the distance between these two points in vowel space.

First, the method used to determine the midpoint between the vowel distributions in perception will be explained. As described in Section 3.1.4, by-subject quasibinomial logistic functions were fitted over the number of /ɛ/ and /o/ responses to determine the boundary width, corresponding to the distance in continuum steps between a 25% and 75% probability of the given response. Boundary width was not chosen as the optimal measure of intraindividual perception variability, given that it does not directly reflect the amount of consistency across identical repetitions. However, the boundary point, which can be calculated from the same statistical models, is a useful measure in assessing participants' categorical boundaries. The boundary point corresponds to the continuum step where a participant had a 50% probability of selecting either of the two options (in this case vowels). This point of perceptual ambiguity is

thought to represent the separation point between two distinct categories (McAllister Byun & Tiede 2017; Park et al. 2019; Cheng et al. 2021). To determine the boundary point, the predicted continuum step value at 50% probability is extracted from each of the subject-specific models built in Section 3.1.4. Specifically for the vowel dimension investigated in this study, the boundary point thus represents the midpoint in F2-F1 space between the distributions of the two vowel pairs (/e-ɛ/ and /u-o/).

This boundary point was associated with the formant values of the corresponding stimulus on the perceptual continuum in the following manner. First, the formant values from the model speaker in the perception stimuli were extracted and processed following the same procedure as in Production B (see Section 2.2.2). Since the participants and the model speaker do not have the same frequency ranges, all participant's formant ranges, along with those of the model speaker, had to be normalized following the method used in Section 2.2.2. To do so, all formant values were z-normalized by subject and these z-scores were then converted back into familiar Hertz values by adding the individual z-score multiplied by the aggregated standard deviation of all 50 individuals (the 49 participants plus the model speaker) to the aggregated mean formant frequency. The two vowel continua of the model speaker, which only consisted of five steps, were then rescaled over 100 points using linear interpolation. As illustrated in Figure 4.9, the perceptual midpoint of each participant was then associated with the closest corresponding formant values in the model speaker's normalized frequency range.

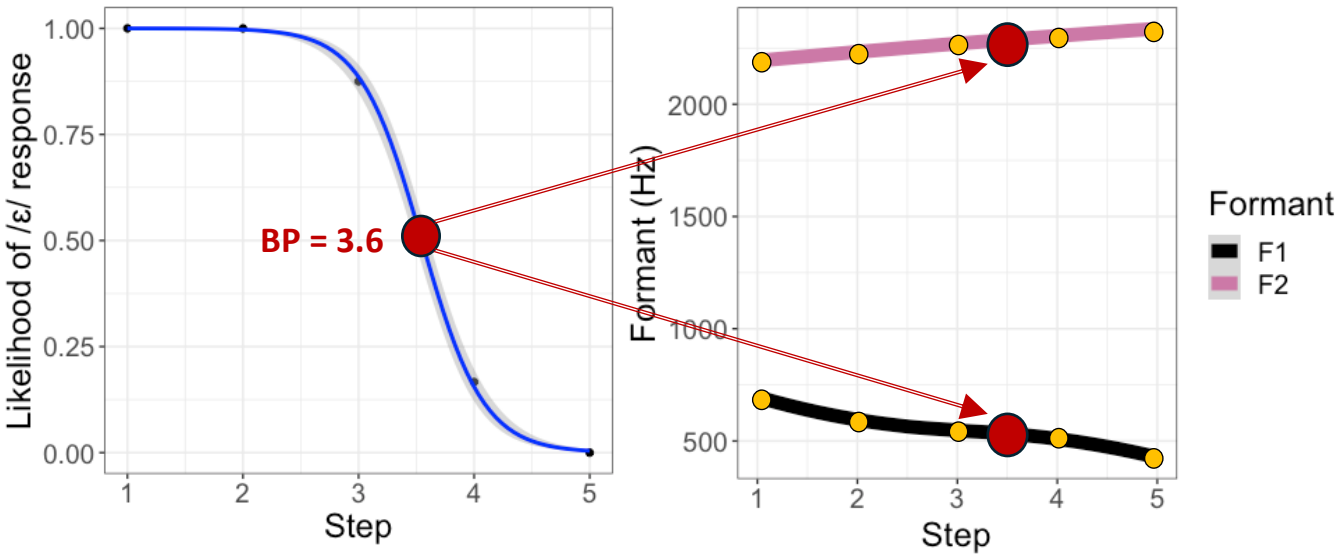


Figure 4.9 Representation of the calculation of the perceptual midpoint for the /e-ɛ/ continuum for a sample participant. First, the boundary point, corresponding to 50% probability of a participant selecting either vowel category, was extracted, as indicated by the red circle in the lefthand plot. Then, the normalized F1 and F2 values of the perception stimulus located at this continuum step were taken as the perceptual midpoint, as indicated by the red circles in the righthand plot. The yellow circles represent the actual formant values of the model speaker extracted from the stimuli used in the perception experiment. Interpolated values were calculated to fill in the spaces between each step for participants whose boundary point did not fall directly on one of the five steps.

As seen in Section 3.2.2, a few participants' response curves for the vowel continua never reached 50%, indicating they consistently chose the same response across most continuum steps and repetitions. As previously discussed, excluding participants simply because their response curves do not follow the traditionally expected categorical curve is misleading (Apfelbaum et al. 2022) and therefore the point closest to the 50% mark was taken as the boundary point for these speakers.

An analogous measure in production was calculated by determining the midpoint between the centroids of an individual's vowel distributions, as illustrated in Figure 4.10. This calculation was inspired by Chao et al. (2019)'s method in determining the boundary between English speakers' productions of the /ɛ-æ/ vowel contrast, although the details differed

considerably. In Figure 4.10, the red triangles represent the centroids of each vowel distribution, corresponding to the mean values of F1 and F2 contained inside the superimposed ellipses. The purple diamond represents the midpoint in F2-F1 space between the centroids of each vowel pair. Essentially, this production midpoint value represents the same construct as the perception boundary point, namely the transition point between two distinct phonological categories.

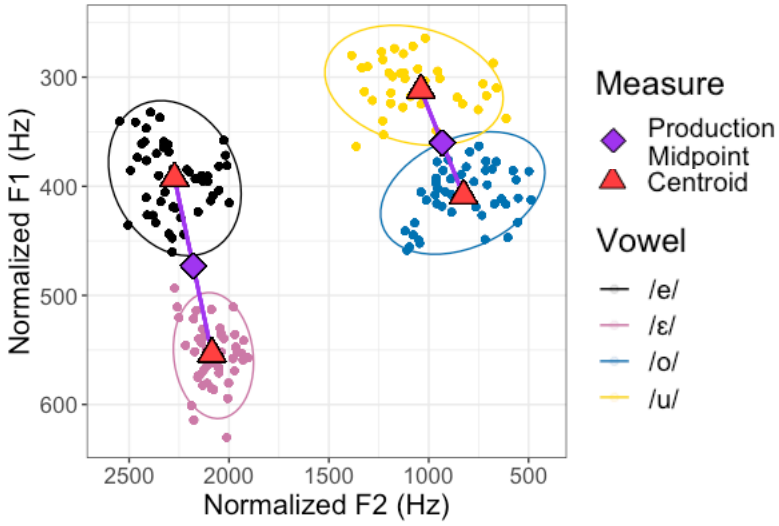


Figure 4.10 Representation of the calculation of the production midpoint for the two vowel pairs (/e-ɛ/ and /u-o/) for a sample participant. Red triangles show the centroid in normalized F2-F1 space of each vowel and purple diamonds represent the production midpoint between the two vowel distributions.

Finally, to calculate the production-perception difference for the two vowel pairs, the Euclidean distance was calculated between the production and perception midpoints for each of the two vowel continua. It is important to note that this production-perception link coefficient is not representative of intraindividual variability, but rather describes the mean category thresholds of each individual.

4.2.2 Results

Figure 4.11 presents individuals' midpoints in production (purple diamonds) and perception (red circles). Immediately apparent is the large degree of interindividual variability. No correlation was observed between production and perception midpoint values for either formant or vowel pair (all p's > .5). This result implies there was no consistent production-perception link at the group level. This differs from the stop voicing continuum where a significant correlation was observed between production and perception cue weights of voicing proportion, as seen in Figure 4.4. This difference is possibly due to the different nature of acoustic cues analyzed. Voicing proportion is largely a durational cue, and listeners essentially would be able to directly map what they hear onto their own values of stop voicing. Therefore, comparing one's production values to those of the model speaker in the perception experiment would not require any real-time normalization, which may possibly facilitate a more direct link. On the other hand, a production-perception link for vowels can only be observed if listeners normalized the formant values of the perception stimuli to match their own vowel spaces, which may require additional effort. The choice of either using such a normalization strategy or taking the formant values as they are without directly comparing them to one's own production values is likely under direct control of an individual. However, it is important to note that while the lack of correlation implies there is no group-level tendency, a strong production-perception link may still be observed for certain individuals.

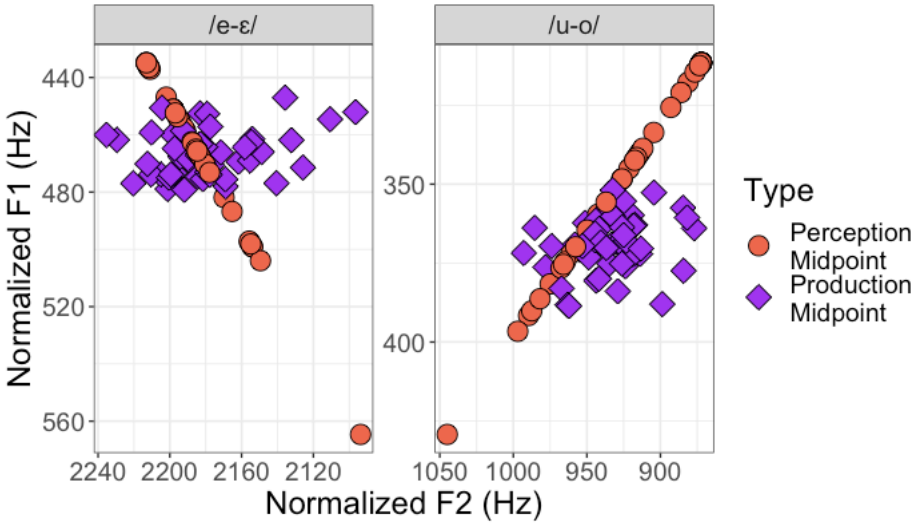
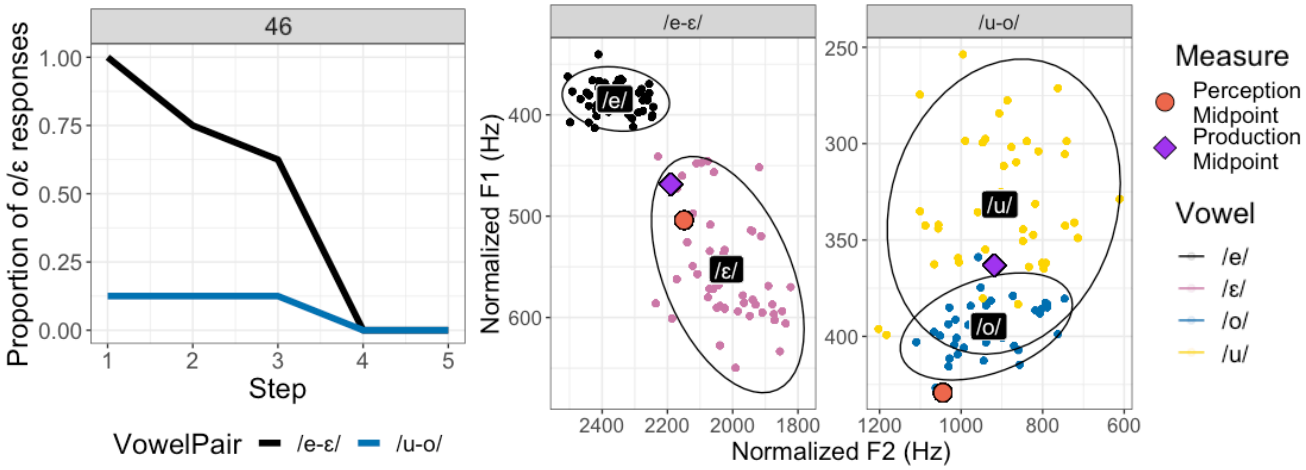


Figure 4.11 Production and perception midpoint values for all participants and both vowel contrasts.

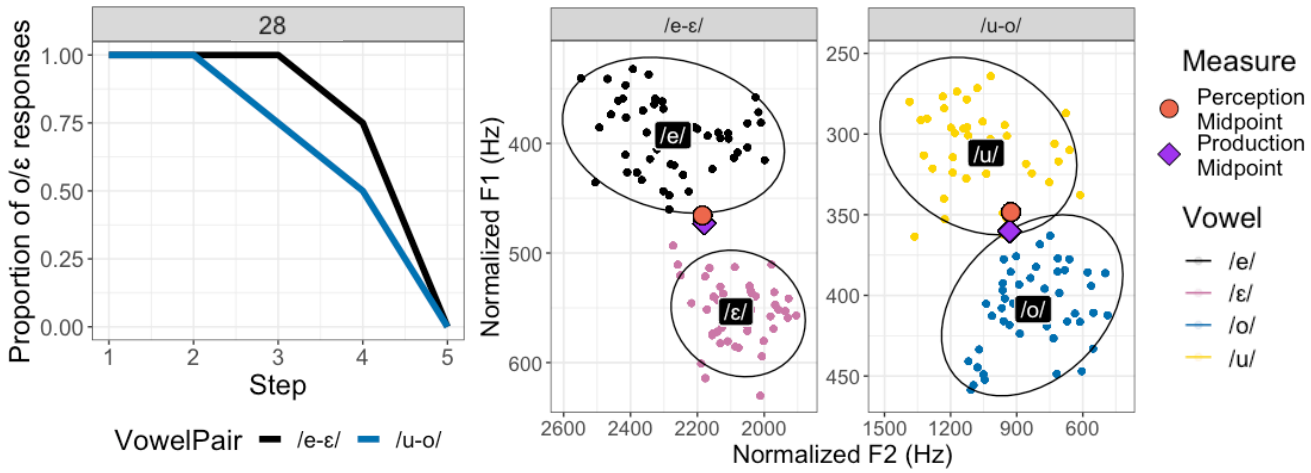
Figures 4.12 and 4.13 present the production and perception data from two participants, exemplifying the accuracy of the vowel production-perception difference measure at capturing the degree of similarity between the two modalities. Participant 46, as can be seen in Figure 4.12, had the highest production-perception difference for the vowel pair /u-o/ and a relatively high difference for the vowel pair /e-ɛ/ as well. This suggests that the speaker's categorization of formant cues in perception does not accurately coincide with the distribution of these same cues in production. This speaker classified /e-ɛ/ with a relatively large boundary width in perception, where the proportion of /ɛ/ responses corresponding to 25%-75% spans three continuum steps. In production, however, the speaker had distinct categories with no overlap. For the /u-o/ continuum, the speaker heard nearly all stimuli as /u/. This behavior is better understood when considering that their formant values for /u/ in production presented considerable overlap with those for /o/. For the /u-o/ continuum at least, it seems that production informed perception in some way, seeing as this individual was possibly less sensitive to the formant differences on the vowel continuum due to the overlapping production categories. However, the midpoint between the two distributions in production was still much higher than the midpoint in perception, which resulted in the large production-perception difference for this vowel pair.



Production-perception difference	
/e-ɛ/ = 54 Hz	/u-o/ = 143 Hz

Figure 4.12 Example participant with a weak production-perception link for both vowel continua. The left plot shows the response to both vowel continua in the perception task. The right plot shows formant values for all tokens in production, as well as the production (purple diamond) and perception (red circle) midpoint, corresponding to the theoretical boundary point between the two phonological categories.

Participant 28, as shown in Figure 4.13, had a strong production-perception link in both vowel categories, evidenced by the very low production-perception difference value. All vowels of this individual had wide distributions yet occupied distinct formant spaces in production. This participant's responses on the /e-ɛ/ continuum were slightly more categorical (evidenced by a small boundary width) than those on the /u-o/ continuum, which is mirrored in production where less overlap was observed between /e/ and /ɛ/ when compared to /u/ and /o/.



Production-perception difference	
/e-ɛ/ = 9 Hz	/u-o/ = 13 Hz

Figure 4.13 Example participant with a strong production-perception link for both vowel continua. The left plot shows the response to both vowel continua in the perception task. The right plot shows formant values for all tokens in production, as well as the production (purple diamond) and perception (red circle) midpoint, corresponding to the theoretical boundary point between the two phonological categories.

The production-perception differences for all individuals and for both vowel pairs is shown Figure 4.14. A lower Euclidean distance is indicative of a tighter production-perception link, where participants' categorization of vowels in perception coincided more with their vowel distributions in production. As can be seen, a tight production-perception link was indeed observed for certain participants, although there is a large degree of interindividual variability.

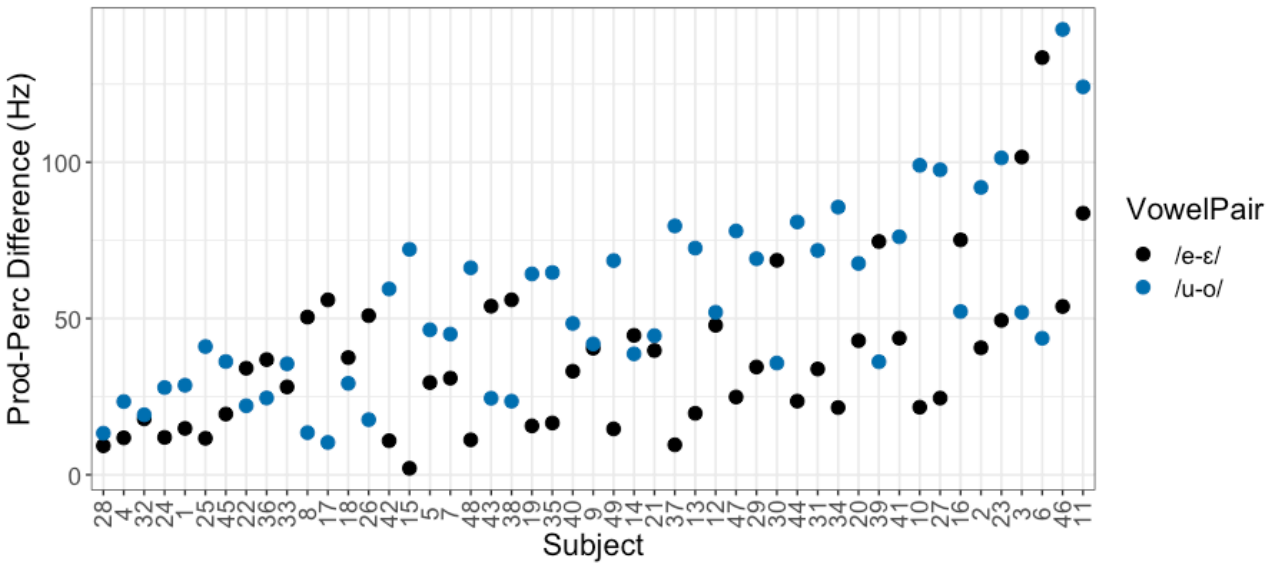


Figure 4.14 Euclidean distances between production and perception boundary points, taken as the production-perception link coefficient. Values are arranged from the lowest (leftmost) to the highest (rightmost) combined sum across vowel pairs.

Overall, the production-perception differences for the vowel pair /e-ɛ/ were slightly smaller than those for /u-o/, as confirmed by a Welch two-sample t-test ($\beta = 17.17$, $t = -3.1$ $p = .003$). Similar to the results in Figure 4.7 for the stop voicing contrast, some speakers present a comparable production-perception difference across the two continua, but no correlation was found between production-perception differences across vowel pairs ($p = .93$). This suggests that the magnitude of an individual's production-perception link is unique to the specific phonological category tested. As observable on the x-axis when comparing Figure 4.14 and Figure 4.7, and confirmed by a Pearson correlation ($p > .1$), an individual with a strong production-perception link for the stop contrast does not necessarily have a strong production-perception link for the vowel contrast, again suggesting that the magnitude of production-perception differences does not represent a global characteristic of an individual. This coincides with the independence of intraindividual variability across phonological categories observed in production and perception. Further correlations exploring the relationship between production-perception differences across phonological categories will be discussed in Section 4.4.

Additional Pearson correlation tests were carried out to determine if the strength of the production-perception link can be predicted by the amount of intraindividual variability present

in either production or perception. A significant correlation for the vowel pair /u-o/ was observed between the production-perception difference and the Coefficient of Unalikeability. However, no significant correlation was observed for the /e-ɛ/ vowel pair, as shown in Figure 4.15. This correlation suggests that participants whose production and perception vowel spaces for /u/ and /o/ matched more closely responded to the perception task with greater response inconsistency, aligning with the result observed for voicing proportion on the stop voicing continuum as shown in Figure 4.8.

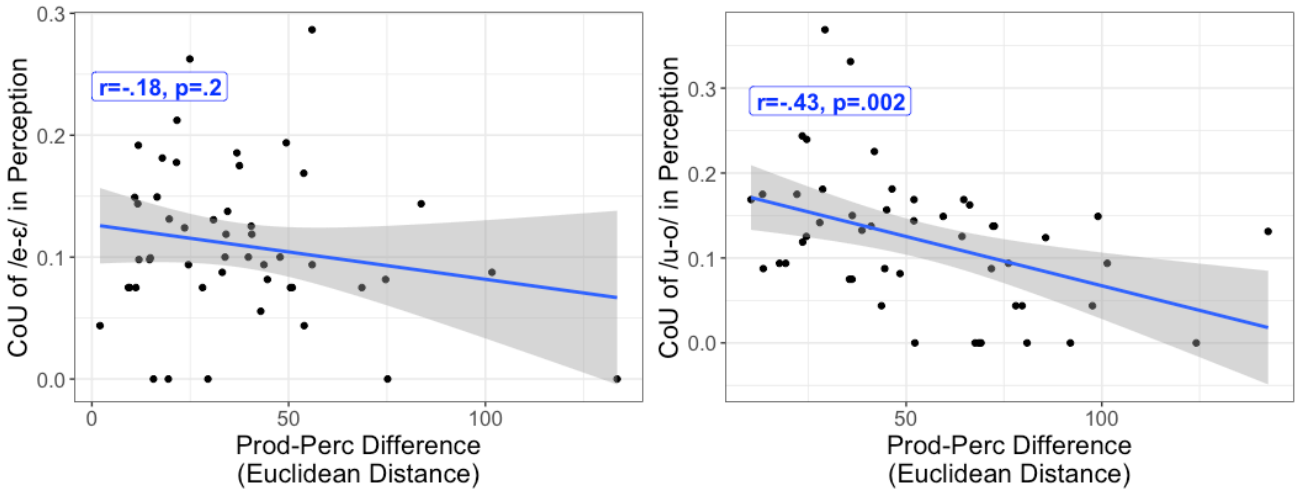


Figure 4.15 Pearson correlations between Coefficients of Unalikeability of the two vowel pairs tested in perception and the Euclidean distance between boundary points in production and perception, representative of the inter-modality link.

As discussed in Section 4.1.2 for the stop voicing continuum, this correlation seems to indicate that greater intraindividual variability on the perception task is an indication of a participant's active use of their own production values, possibly due to real-time comparison of what they hear with what they produce themselves. This strategy of accessing one's production values anew even on identical trials in a perception task may be most necessary when hearing more ambiguous stimuli. If this is the case, the lack of correlation for the /e-ɛ/ vowel pair could indicate that the stimuli on this continuum were less ambiguous than the other continua and that their categorization thus required less effort. Indeed, less intraindividual variability was observed for the /e-ɛ/ vowel continuum than for the /u-o/ continuum, as shown in the CoU values in Figure

3.12, suggesting participants responded with greater certainty to the /e-ɛ/ stimuli. In addition, the fact that certain participants exclusively chose /u/ on the /u-o/ continuum indicates that the /u-o/ identification was likely more difficult. This possibility will be further explored in Section 4.5 when the influence of reaction time on the production-perception difference is considered.

It is interesting to note that participants whose unnormalized formant values were close to those of the model speaker in the perception experiment, which would arguably reduce the need for inline normalization to compare with one's own production values, did not necessarily have a stronger production-perception link. This was confirmed by the lack of a significant Pearson correlation between the production-perception difference coefficients and the average formant differences between the model speaker and each participant ($p = .8$). This suggests that the raw distance between one's own formant ranges and those of one's interlocutor does not directly affect the presence or strength of a production-perception link.

Regarding the effect that production variability may have on the strength of a production-perception link, no significant correlation was observed between the production-perception difference and the CoV of either F1 or F2 for any vowel (all $p > 0.2$). The strength of the production-perception link, as measured here, is thus mostly dependent on participants' performance in the perception experiment. This offers preliminary evidence that production and perception may be connected under certain circumstances but suggests that this link is not automatic and is likely under the listener's control to a certain degree. This will be further discussed in Section 4.6 and Chapter 5.

4.3 Fricatives

4.3.1 Methodology

The same method as described in Section 4.2.1 for vowels was employed to determine the production-perception difference of the fricative contrast. First, individual perception boundary points between /s/ and /ʃ/ were extracted from by-subject quasibinomial logistic functions. These boundary points were then associated with COG values as follows. The COG values of the 49 speakers and of the model speaker were z-scaled and then converted back to

Hertz values following the methodology described in Section 4.2.1. The perception boundary point of each participant was then associated with the corresponding normalized COG value to obtain the perceptual midpoint between the two fricative categories. The categorical boundary in production, on the other hand, corresponded to the midpoint between the means of normalized COG values of /s/ and /ʃ/. The absolute value of the difference between the production and perception midpoint was taken as the production-perception link coefficient.

4.3.2 Results

Coinciding with the result for the vowel pairs, no significant group-level correlation was observed between the categorical boundary points in production and perception of the fricative contrast ($p = .5$), again indicating that the presence of a production-perception link is not an automatic process. Figure 4.16 presents the production-perception differences of all participants. Immediately apparent is the high amount of very large production-perception differences, indicative of a very weak link. These participants correspond to the groups Only /ʃ/ and Only /s/, as established in Section 3.2.3, who chose either /ʃ/ or /s/, respectively, at least 95% of the time. As a reminder, this was interpreted as a clear sign that listeners employ distinct strategies when carrying out the same task, and that at the individual level, a single participant's performance varies even across similar tasks. In other words, participants reacted differently to the categorization of distinct phonological segments even within the same experimental session.

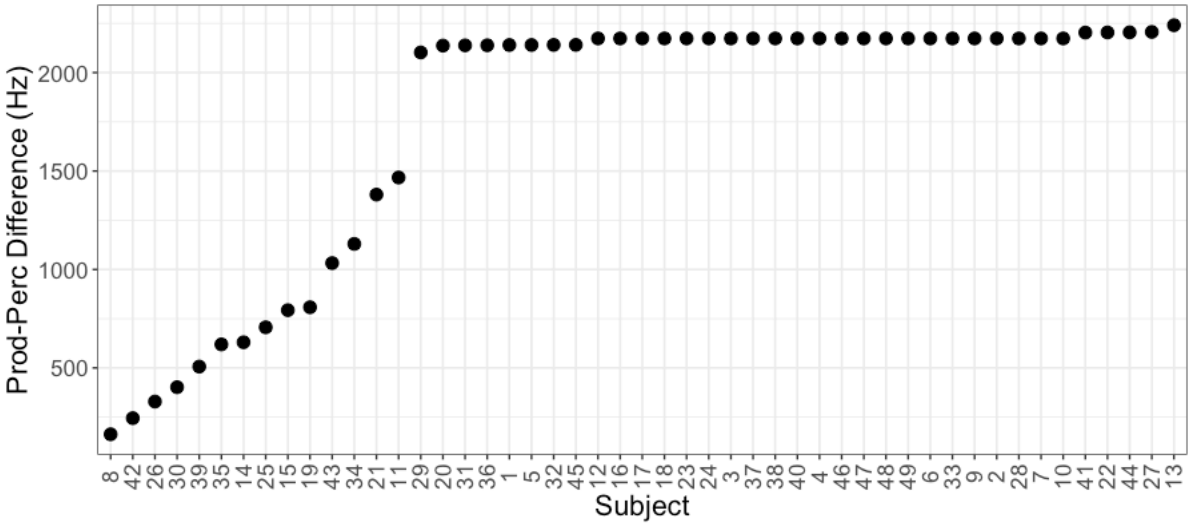


Figure 4.16 Production-perception link coefficients for the /s-ʃ/ fricative contrast, arranged from lowest (leftmost) to highest (rightmost).

The production-perception difference for fricatives was not significantly correlated with intraindividual variability in production (CoV of COG) of either /s/ or /ʃ/ ($p = .5$). Therefore, the strength of the production-perception link is not directly affected by participants' production tendencies, as was observed for the vowel contrasts. However, a significant negative correlation was observed between the production-perception difference and the Coefficient of Unalikeability, representing response inconsistency in the perception experiment, as shown in Figure 4.17. When considering the three groups observed in perception, as described in Section 3.2.3 (namely those that heard both /s/ and /ʃ/, those that chose /s/ 95% of the time and those that chose /ʃ/ 95% of the time), the negative correlation was only found in the group that chose both fricatives, although it no longer reached significance ($r = -.4$, $p = .09$).

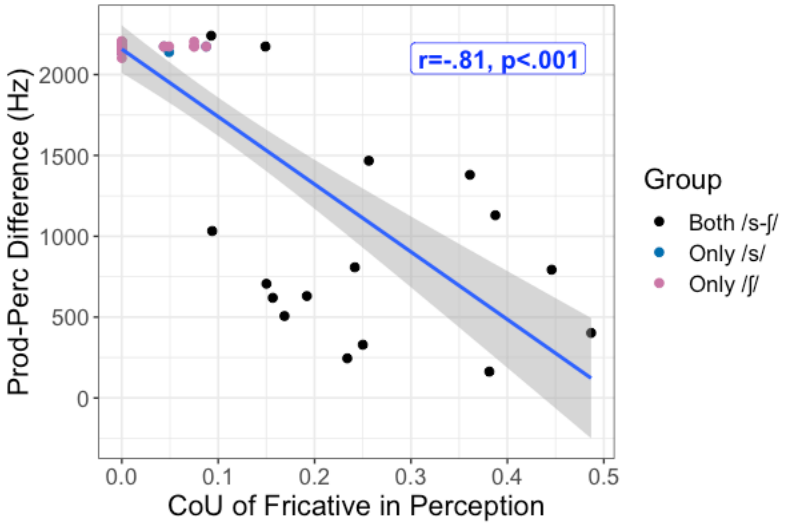


Figure 4.17 Pearson correlation between the Coefficient of Unalikeability of fricative pair /s-ʃ/ in production and the fricative production-perception link coefficient. Solely for visualization purpose, the three participant groupings for this continuum, as explained in Section 3.2.3, are shown.

This result aligns with the similar significant correlations for voicing proportion on the stop continuum and the /u-o/ vowel continuum, indicating that participants who exhibited greater variability in the perception task displayed a tighter link between their acoustic ranges in production and perception. However, for the fricative continuum, a CoU of 0, while still representing complete response consistency, is less informative than for the other continua given that most individuals with a CoU of 0 exclusively selected one of the two fricatives on all trials. Nevertheless, it is intriguing to notice that within the group of Both /s-ʃ/ who made ample use of both categories, as response inconsistency increased, participants' responses more closely matched their production values. Overall, this result, in particular for the fricative continuum, indicates that variability is likely not a sign of a lack of attentiveness, but rather a reflection of a certain approach to the phoneme categorization task, namely one where individuals actively consult their own production values upon hearing each token, regardless of if they have heard this token previously.

4.4 Link between variability

Given the existence of a relatively strong production-perception link for some participants, the question arises if the degree of the production-perception difference is similar across segments and/or acoustic properties. No Pearson correlation between the five production-perception difference measures was significant ($p > .2$), as shown in Table 4.1. This implies that an individual with a tight link for one acoustic dimension or phonological segment does not necessarily have a tight link for other acoustic dimensions or phonological segments. Specifically, no correlation was observed even between the two vowel continua (/e-ɛ/ and /u-o/) nor across the two voicing cues manipulated on the stop continuum (voicing proportion and post-release aspiration duration).

Table 4.1 Pearson correlation coefficients between production-perception differences for the five categories tested. No correlation was significant ($p > .2$).

	Stop voicing	**Aspiration**	**/e-ɛ/**	**/u-o/**	**/s-ʃ/**
Stop voicing		r = .19	r = .07	r = -.19	r = .06
Aspiration	r = .19		r = -.11	r = .03	r = .14
/e-ɛ/	r = .07	r = -.11		r = -.01	r = -.03
/u-o/	r = -.19	r = .03	r = -.01		r = .16
/s-ʃ/	r = .06	r = .14	r = -.03	r = .16	

Additionally, Pearson correlations were carried out between CoV and CoU values to evaluate if a high degree of intraindividual variability in one modality may affect the degree of variability in the other modality. To align the variability measures in production with the five continua used in perception, average CoV values were used, similar to those shown in Figure 2.25. Specifically, two CoV values were calculated for stops, one for voicing proportion and one for post-release aspiration duration, averaging across place of articulation. For vowels, the CoV of F1 for the vowels /e/ and /ɛ/ were averaged together to obtain a single /e-ɛ/ CoV of F1 value and the CoV of F1 for the vowels /u/ and /o/ were combined to obtain a single /u-o/ CoV of F1 value.

The same was done for the CoV of F2 for both vowel pairs. Finally, a single CoV of COG value was calculated by averaging the values for /s/ and /ʃ/. The CoU for the stop voicing continuum corresponded to the average value across the four stop voicing continua, shown previously in Figure 3.15. As shown in Table 4.2, no correlation was significant, indicating that the amount of intraindividual variability for a given phonological segment or acoustic property is specific to the modality measured. In other words, even though those individuals who were found to have a strong link between production and perception likely made use of their own production values during the perception task, suggesting some type of connection between acoustic values in both modalities, their rates of variability in each of the modalities were independent from one another.

Table 4.2 Pearson correlation coefficients and p-values between Coefficients of Unalikeability (CoU) values representing intraindividual variability in perception and Coefficients of Variation (CoV) values representing intraindividual variability in production.

Contrast	**r**	**p**
CoU of **Stops** – CoV of **Voicing proportion**	.25	.08
CoU of **Stops** – CoV of **Aspiration duration**	.25	.08
CoU of **/e-ɛ/** – CoV of **F1**	-.08	.6
CoU of **/e-ɛ/** – CoV of **F2**	-.03	.8
CoU of **/u-o/** – CoV of **F1**	.04	.8
CoU of **/u-o/** – CoV of **F2**	.03	.8
CoU of **Fricatives** – CoV of **COG**	-.02	.9

The lack of correlation between variability in production and perception within a phonological category suggests that intraindividual variability is not only specific to articulatory configurations and their corresponding acoustic properties but also to each speech modality. Crucially, this means that intraindividual variability is not a generalizable speaker-specific trait. These results are in line with Cheng et al. (2021) who observed no correlation between intraindividual variability of English vowels in production and perception. This differs, however, from Chao et al. (2019)'s observation that an individual's categorical boundary in perception is

further away from the more variable vowel in their production, which was taken as evidence that production and perception variability are correlated. The methodological differences between Chao et al. (2019)'s study and the current dissertation make it difficult to directly compare results. Not only did Chao et al. (2019) operationalize intraindividual variability in a different manner, but the participants in their study classified perception stimuli derived from their own voices. Cheng et al. (2021), who used a model speaker in the perception experiment as was done in the current dissertation, speculated that the discrepancy between these results is because a strong link between variability in production and perception may only be present when listening to one's own voice. Given that proprioception is known to be an important part of language acquisition (Meltzoff 1993), this is a plausible explanation. In addition, perceiving one's own voice and perceiving sounds produced by others are likely two distinct cognitive tasks. Overall, the results of this dissertation coincide well with Cheng et al. (2021)'s results: intraindividual variability is not a generalizable subject-level characteristic but is rather unique to specific tasks. In production, these specific tasks correspond to distinct articulatory configurations while in perception these tasks equate to the categorization of distinct aspects of the acoustic signal. Given that production and perception themselves are highly different mechanisms that make use of different anatomical structures, it is unsurprising that intraindividual variability is not linked across these two modalities.

4.5 Influence of participant characteristics

In light of the observation in Section 3.2.3 that the age of acquisition of L2 moderately predicted the amount of variability present on the /u-o/ and /s-ʃ/ continua of the perception experiment, it is pertinent to consider the effect this and other sociodemographic and linguistic background-related factors may have on the strength of the production-perception link. The factors considered, previously presented in Table 3.5, are repeated here in Table 4.3. Independent Welch t-tests were used to analyze the categorical variables and Pearson correlation analyses were used to analyze the numerical variables, as specified in Table 4.3. Analyses of possible interactions were not considered as this led to highly uneven distributions

across groups. Dependent variables were the five production-perception difference measures presented above (voicing proportion of stops, post-release aspiration duration of stops, vowel contrast /e-ɛ/, vowel contrast /u-o/, fricative contrast /s-ʃ/).

Table 4.3 Levels and counts of sociodemographic and linguistic variables. Categorical variables are written in italics and values in parentheses represent subject counts. Numerical variables are marked in bold with ranges are given in parentheses.

Variable	Levels
Native bilingual (Laurentian French-English)	*Yes (16) – No (33)*
Province of origin	*Ontario (23) – Quebec (26)*
English proficiency	**5 values (native – low)**
Age of acquisition of L2	**14 values (3 – 18)**
Age	**11 values (18 – 45)**
Gender	*Female (35) – Male (13) – Non-binary (1)*
Presence of L3	*Yes (20) – No (29)*
Proficiency of L3 (n=20)	**4 values (high – low)**
Time spent outside Ontario/Quebec	*0 years (45) – 1-2 years (2) – 3-4 years (2)*
Percentage of daily use of Laurentian French	*80-100% (24) – 50-80% (20) – 20-50% (5)*

No sociodemographic or linguistic factor shown in Table 4.3 significantly predicted or was significantly correlated with any of the production-perception difference measures. It is intriguing that the amount of intraindividual variability in production, perception and the cross-modality link are largely undetermined by sociodemographic or linguistic factors, apart from moderate effects for native bilingual status and age of acquisition of L2, as observed in Section 3.2.3. The lack of such significant tendencies bolsters the conclusion that variability is a unique characteristic of the individual and further highlights the importance of considering how such individual variability may influence broader observed tendencies in linguistic experiments. It also raises the question of why such variability exists and whether individuals who produce or

perceive certain phonological or phonetic segments with greater variability obtain a linguistic advantage or disadvantage. These questions will be further explored in Chapter 5.

As described in Section 3.2.4 and shown in Figure 3.17, longer mean reaction times were found to correlate with greater degrees of intraindividual variability in perception (measured by CoU). It was suggested that this is likely indicative of higher attention in the identification task. However, longer reaction times can just as well be an indication of boredom or distraction which would generally lead to less informative participant responses. To tease apart these two interpretations, Pearson correlations were carried out between mean reaction times for each of the four continua (the four individual stop voicing continua grouped together, the two vowel continua and the fricative continuum) and the production-perception link coefficients. As shown in Figure 4.18, significant correlations were observed for the two vowel continua and the fricative continuum, suggesting that participants who more actively compare their own production values with the perception stimuli take longer to reach a decision.

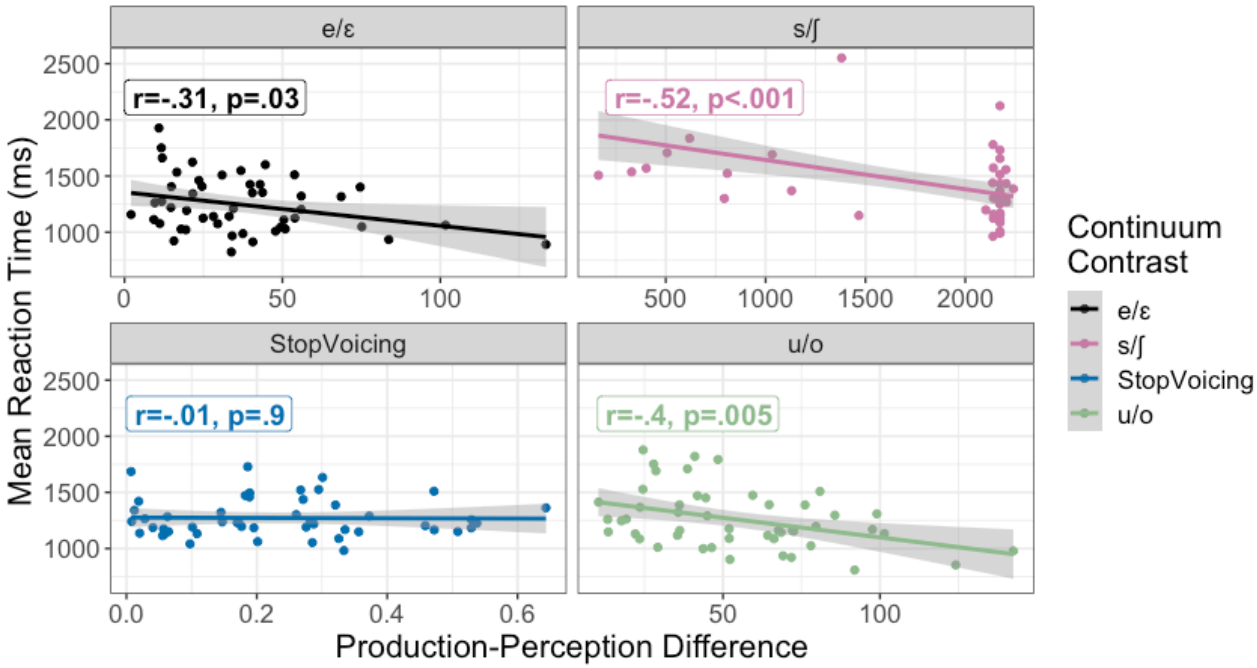


Figure 4.18 Pearson correlations between mean reaction times and production-perception differences for the four perception continua measured. The production-perception difference for the stop voicing continua represents only voicing proportion.

Neither the production-perception difference for voicing proportion nor for burst duration was correlated with the mean reaction time for the stop voicing continua (only the correlation between the mean reaction time and the production-perception difference for the voicing proportion cue are shown in Figure 4.18). This is likely due to the differences in task type across the continua. As a reminder, the vowel and fricative continua probed the location of participants' categorical boundaries while the stop voicing continua assessed the weight of two acoustic cues. In addition, the vowel and fricative continua required perception of spectral characteristics which vary considerably from one individual to another, forcing individuals to carry out some sort of normalization with the model speaker's speech. This is likely not necessary for the stop voicing continuum, given that both voicing proportion and post-release aspiration duration are largely durational cues. It may be that the additional step of normalization for the vowel and fricative continua increased the temporal demands for those participants who chose to compare stimuli with their own production values.

The fact that, overall, participants who took longer to reach a decision chose values in perception closer to their values in production offers convincing evidence that longer reaction times are indeed indicative of more attentiveness, at least on this 2AFC perceptual categorization task. This aligns well with the significant correlations observed between CoU and the production-perception difference, presented in this chapter, and the significant correlations between CoU values and mean reaction times, discussed in Section 3.2.4 and shown in Figure 3.17. In other words, longer reaction times are associated with greater variability due to an active comparison with one's own production values upon hearing each stimulus, regardless of how one previously categorized the same stimulus. Individuals likely analyzed each stimulus as a new linguistic object, possibly sampling at different temporal moments or changing the weight they assigned to different aspects of the speech signal. Particularly on difficult tasks, such changes of mind in decision-making even across similar trials are often observed (Resulaj et al. 2009). It is important to note that the moderate nature of the correlations in Figure 4.18 points to a complex picture, where, again, individual differences are greater than any immutable, subject-independent tendency. However, given the presence of this correlation across three perceptual continua, the

proposal that greater intraindividual variability in the perception task is an indication of an active use of production cues in stimuli categorization is supported.

4.6 Discussion and conclusion

Overall, the strength of the production-perception link for all four acoustic cues measured, i.e. stop voicing proportion, stop post-release aspiration duration, vocalic formants and fricative COG, differed greatly among individuals. A group-level correlation between the two modalities was not observed for the vowel or fricative contrasts but was found for the stop voicing contrast, though considerable interindividual variability remained. Additionally, the strength of the production-perception link was not correlated across distinct phonological segments within the speech of an individual. These results coincide with the observations regarding intraindividual variability in production and perception, where the degree of variability was dependent on the precise phonological segment or acoustic cue measured, rather than corresponding to an overall individual trait.

As summarized in Section 1.2.2 and Section 1.2.3, a plethora of studies have investigated possible production-perception links, both for cues participating in sound changes, as well as cues thought to be more stable, and in both cases, strong evidence for and against a direct link has been found. The observations made in this dissertation support the idea that the strength of such a link is highly dependent on the individual and the segment analyzed, which could explain the lack of consistent observations in previous studies. There seem to be two possible interpretations. The presence or absence of significant group-level tendencies in previous studies is merely a byproduct of the random sampling of participants, which arbitrarily included more or less individuals with strong links. Conversely, such a production-perception link may be under an individual's control in some way and the tasks commonly used to test perception in linguistic experiments may allow one to choose to make use of this connection or not. Regardless, the lack of a strong and consistently present production-perception link is a somewhat expected conclusion, given that for effective communication, individuals naturally must consider a range

of production patterns in perception, even if they themselves do not produce these patterns at all.

Greater response inconsistency in the perception task was found to be correlated with a tighter link between production and perception for the voicing proportion cue in stops, the /u-o/ vowel continuum and the fricative continuum. These correlations, coupled with the observation that longer mean reaction times were significantly correlated with tighter production-perception links, implies that greater intraindividual variability on a perception task is representative of an active use of production cues and that this strategy requires more time. Upon hearing each stimulus, the individual carries out a sort of probability calculation based on their own production categories, which may lead to distinct responses across trials. However, greater response inconsistency could also simply be a consequence of individuals with a greater tolerance for phonetic variation, as has been suggested in studies probing the presence of *categorical* and *gradient* perceivers (e.g., Kapnoula et al. 2017, Honda et al. 2024a). Rizzi & Bidelman (2024) observed that more gradient listeners on both a 2AFC and a VAS task had longer reaction times and stronger neural activations, which they also interpreted as evidence of greater attention and increased task engagement. Given the significant correlation presented in Figure 4.18, the longer reaction times Rizzi & Bidelman (2024) observed in gradient listeners may indicate that they compare perception stimuli more often with their own production categories. That said, the fact that the observed correlations are moderate and were not found for all categories also suggests that there is likely more to the picture than what has been captured by the data here.

It should be recognized, however, that individuals with greater production-perception differences, and thus generally lower rates of response inconsistency, do not necessarily always have a weak production-perception link. The absence of such a link in this experiment may merely reflect a task-specific strategy where participants chose to not make active use of their own production targets when categorizing stimuli. Indeed, it is important to avoid the assumption that all participants will employ the same response strategy when carrying out a repetitive and artificial tasks, such as the 2AFC identification task used here. Given the design of the task, where listeners are presented with ambiguous, often odd-sounding stimuli, as well as the absence of

the rich linguistic or extralinguistic context that is normally present in real-life speech perception, some individuals may base their decision on specific aspects of the acoustic signal rather than interpret the stimuli as language objects.

Clearly, the presence of correlations between the production-perception difference and both reaction times and rates of response inconsistency in the perception experiment reveal that some individuals accessed their production cues during perception. However, both reaction times and response inconsistency measures come from the perception experiment itself and more importantly, intraindividual variability measures obtained from the production experiment were not correlated with the production-perception differences. This suggests that the presence of a production-perception link may be more indicative of a strategy employed specifically to complete the 2AFC task, rather than a construct used in real-life situations. In order to find out if the production-perception link is more than just a specific response strategy to the perception task, it would be necessary to show that it is bidirectional, i.e., that participants reference perception cues in some way when producing speech. This is unfortunately impossible to determine with the data in the current dissertation. Although in real instances of language use listeners may at times actively weigh acoustic information to assign an ambiguous sound to a phonological category, context cues (syntactic, intonational, discourse, etc.) are likely almost always sufficient to ensure accurate perception in native speakers. Therefore, it remains unclear if there would ever be an actual need to actively access one's own production cues in the perception of spontaneous, natural speech. Regardless, the most pertinent conclusion of this chapter is that production-perception differences between acoustic cue weights and phonological categories, just like intraindividual variability in each modality, is unique to the individual and to the segment analyzed. This specificity implies that production and perception of different linguistic segments represent distinct tasks and that individuals can differ in their physical capacities and tolerance levels across such linguistic tasks. The ubiquity, structure, and possible advantages of intraindividual variability in speech, and more broadly across different domains of human behavior, will be discussed in Chapter 5.

5. Discussion

The main results of the analyses presented in the preceding chapters can be summarized in the three following conclusions:

1. The degree of intraindividual variability in production and perception is similar for segments that share a common acoustic target and/or articulatory configuration.
2. Intraindividual variability of any given segment or acoustic property is not correlated between production and perception.
3. The strength of the production-perception link in acoustic cue use differs across individuals as well as within individuals across segments and increases in subjects with higher response inconsistency in the perception experiment.

In this chapter, these conclusions will be described in greater detail and compared with findings from previous studies in both linguistics and other domains that probe intraindividual variability. Then, how the results of this dissertation can inform current models of production and perception will be briefly examined, followed by a discussion of the role of variability and of the possible advantages that could explain its pervasiveness.

5.1 Individual variability in production

Overall, the degree of intraindividual variability in production was found to be correlated across segments with similar acoustic values. Rates of variability in voicing proportion and post-release aspiration duration were similar in voiced and voiceless stops, respectively, across distinct places of articulation. Likewise, vowels produced with similar F1 and F2 values exhibited comparable rates of variability. No correlation was observed in the variability of COG between /s/ and /ʃ/, despite a shared status as coronal fricatives, as their COG values occupy a distinct frequency range.

Segments with similar acoustic values are normally produced with similar articulatory gestures. For example, voiced stops, regardless of place of articulation, require a highly similar configuration of the larynx to produce and sustain vocal fold vibrations (Ohala 2011). Along the same lines, voiceless stops, regardless of place of articulation, require an open glottis during the closure and a similar control of vocal fold adduction upon stop release. Distinct values of F1 and F2 are also largely determined by different tongue positions that can be shared across vowels. On the other hand, despite a common identity as coronal fricatives, both /s/ and /ʃ/ are produced with distinct tongue positions and shapes (Toda 2009).

While a change in a single articulator has been shown to have a somewhat predictable effect on the resulting acoustic signal, there is not always a one-to-one correspondence between articulation and acoustics, precisely because the specific characteristics of the speech signal are affected by the configuration of the entire articulatory apparatus (Iskarous et al. 2011). Given this caveat, and the lack of articulatory data in the current dissertation, the question of which articulators are most responsible for variability cannot be clarified here. In other words, it remains undetermined if the correlation in variability between F1 values of the vowels /e/ and /o/, for example, is due to the overall tongue shape (which is arguably considerably different for the two vowels) or specifically to the very similar position of the tongue body (high in the oral cavity). Despite this limitation, it is more likely that similar degrees of intraindividual variability are correlated across specific regions of articulatory space, rather than the entire phonatory apparatus. Harper (2021) tackled this very question with articulatory data and observed few correlations between the intraindividual variability of distinct articulatory points (tongue body, tongue tip, lip aperture, etc.) even within the same segment. These results suggest that similar rates of variability are indeed a property of a specific articulator or articulatory region. It therefore can be postulated that the correlations in intraindividual variability observed between segments are likely due to a high degree of similarity in the articulatory configuration that most affects the acoustic dimension analyzed. In addition, the strength of these correlations generally coincides with the degree of similarity between articulatory gestures. For example, correlations between the CoV of voicing proportion across places of articulation was very high (all $r > .64$, see Figure 2.13), in all likelihood because the larynx configuration is nearly the same for all voiced

stops. The correlation between the CoV of F1 in /o/ and /e/ was also very high (r = .83, see Figure 2.19), which can be explained by the fact that both vowels have a nearly identical F1 value and are produced with highly similar tongue heights. The correlation between CoV of F2 in /ɛ/ and /a/, however, was moderate and less significant (r = 0.32, p = .02, see Figure 2.26), given that despite certain similarities in horizontal tongue position and F2 values, there remain non-negligeable differences between the two vowels in acoustics and articulation.

Regarding the magnitude of intraindividual variability for specific segments, overall higher rates were observed for post-release aspiration duration and F2. Nevertheless, a large amount of interindividual variability was still observed. For example, whereas some speakers' variability of F2 was greater than their variability of voicing proportion in stops, other speakers showed the opposite tendency (see Figure 2.25). Finally, the analysis of the ten speakers who carried out the production experiment twice revealed that the magnitude of variability within an individual segment is a stable property of individual speakers, as no significant differences between the two recording sessions were observed (see Section 2.4). Overall, the results from the analysis of production variability show that the precise structure of intraindividual variability is a unique characteristic of each language user and can be likened to an individual "fingerprint".

Altogether, although the conclusion regarding the structure of intraindividual variability would be bolstered by more direct articulatory data, the observed tendencies offer strong support that speech variability is articulatory-based. Speech articulation involves the precise coordination of many distinct muscles, ranging from the fine-tuned control of respiration, vocal fold configuration or tongue and lip position, as well as more ancillary features such as soft tissue tension. Given that speech articulation is rooted in motor control, a feature which underlies a plethora of additional tasks inherent to the human experience, such as locomotion, food intake, tool use, etc., it is paramount to examine the presence and role of intraindividual variability on a larger scale, considering language-external aspects of human motor abilities.

The comparison of variability in speech production with other aspects of motor control is only justified if both aspects of human behavior are indeed controlled by shared structures in the brain. Although an in-depth review of research into neural activation in speech articulation lies outside the scope of this dissertation, the findings of such studies provide support for the claim

that speech is but one aspect of a shared motor system. Many neuroimaging studies have identified brain regions and neural activity patterns specifically associated with the articulation of speech (Bohland & Guenther 2006; Basilakos et al. 2018). Bouchard et al. (2013), for example, found that, when three English-speaking subjects produced the phonemes /d, b, g, i, u, a/, distinct regions of the ventral pre- and post-central gyri were activated, offering evidence that specific lingual positions cause activation of distinct sites. In a similar manner, Mugler et al. (2018) found that the cortical representation of speech sounds in English speakers, obtained through electrocorticography, corresponds more to gestures, i.e. specific articulators, than to phonemes. Specifically, Mugler et al. (2018) observed that the activity patterns of one allophone of a given phoneme were less similar to other allophones of this phoneme than to different phonemes produced with similar gestures. For example, in the cluster /st/, a /t/ produced with a short duration of post-release aspiration patterned more with word-initial /d/, which is generally also realized with a lack of closure voicing and a short duration of post-release aspiration, than word-initial /t/, which is produced with a much longer duration of post-release aspiration. In other words, speech production representations in the brain appear to be more aligned with articulation rather than phonemic classification.

Despite the identification of such speech-specific brain regions and neural activity patterns, there is also a substantial amount of overlap between speech-related and other movement-related neural dynamics. Regions of the motor cortex associated with control of orofacial and laryngeal muscles necessary for speech have been shown to also be activated during the realization of non-speech tasks, such as production of non-speech onomatopoeic sounds like imitating wind or a car (Lancheros et al. 2020), respiration (Basilakos et al. 2018), and coughing or kissing (Chang et al. 2009). Nevertheless, the extent and amplitude of activation is often higher in non-speech compared to speech, possibly due to the more novel nature of the former (Chang et al. 2009). Additionally, fMRI studies have shown that upon producing speech, certain motor areas corresponding to hands and legs are activated concurrently, implying that speech articulation, at least in part, relies on similar brain regions and motor control structures as those involved in non-speech movements. For example, Liuzzi et al. (2008) observed that reading out loud increased the corticospinal excitability of leg muscle representations. Vainio et

al. (2013) observed increased precision in precision grip tasks, i.e. a task involving only the fingers, when participants produced the syllabe [ti] and increased precision in a power grip task, i.e. a task requiring the use of the whole hand, upon hearing the syllable [ka]. Their interpretation of this result was that the precision grip task can be seen as a counterpart to the high precision necessary in tongue tip placement and small mouth aperture in [ti] production. On the other hand, the use of a larger mouth aperture and greater surface area of the tongue for [ka] parallels the magnitude of a whole hand movement. Komeilipoor et al. (2016) confirmed that the hand muscles involved in these motor tasks were indeed activated when participants produced either of the aforementioned syllables, even when not simultaneously performing these motor tasks. Such results suggest that muscles involved in distinct tasks are not necessarily separated modules in the brain, but rather belong to a common motor control system and coactivation is a common occurrence.

Given the observation of overlapping activation in motor control of speech, non-speech sounds, and other movement types which are thought to be largely unrelated to the articulation of speech, such as that of limbs, it is highly pertinent to consider the possible parallels between motor variability in non-linguistic, movement tasks and the speech variability observed in this dissertation. Indeed, motor variability is a well-researched topic, due to its implications in understanding rates of injury risk, ease of motor learning and consequences of fatigue in repetitive tasks (Getchell et al. 2001; Oomen et al. 2023; Lyons et al. 2025). In studies investigating motor variability, participants are generally asked to complete specific tasks that can be simple, such as pointing, clapping, or finger tapping, or more complex, such as lifting an object, placing an object on a shelf, or galloping. Motor variability is captured through a wide range of techniques, largely determined by the nature of the task: electrodes placed on hands or feet to measure frequency or force of clapping or walking (Getchell et al. 2001; Streamer & Queen 2024), interactive objects measuring keypresses for tapping tasks (Getchell et al. 2001), the force in a pinching task (Lyons et al. 2025), or motion capture data with markers placed on distinct parts of the body (Yang et al. 2018; Oomen et al. 2023).

Most studies on motor variability have determined that variability is not a generalizable, individual characteristic, but rather is dependent on the specific task and the muscles needed to

perform each task. Getchell et al. (2001) measured the coordination frequency of participants performing a wide range of tasks, such as finger tapping, crawling, galloping, clapping, and clapping while galloping. Variability was measured by means of the Coefficient of Variation (CoV). Neither the overall CoV across tasks nor the CoV of the same limb across different tasks was correlated, emphasizing that the rate of variability is task-specific and not dependent on the group of muscles activated, but rather the precise configuration and employment of these muscles. These results align well with the conclusions of the current dissertation, where variability is not a consistent, individual characteristic, but rather highly dependent on the specific articulatory configuration underlying the measured acoustic property. In another study, Haar et al. (2017) also measured the Coefficient of Variation of arm direction, movement extent and peak velocity in supine individuals carrying out a reaching task using a single arm. Participants carried out the reaching task by performing slicing movements with a pen on a digital tablet to four distinct targets located at distances of 7 and 13 cm at a ±45° angle from the starting point. They found a strong correlation between all kinematic measures of the same arm to distinct targets and all measures between the right arm and the left arm to the same targets. The correlations across similar tasks to distinct but close targets mirrors the strong correlation of intraindividual variability of vowels sharing a common space in the oral cavity (such as /u-y/ or /o-e/) that was observed in Production B.

Other studies quantify movement variability by means of Standard Deviation (Kasuga et al. 2022; Oomen et al. 2023). Kasuga et al. (2022) observed movement variability in 25 individuals performing a range of tasks involving hand-arm coordination, such as reaching towards a target or hitting a ball with a virtual paddle. Participants' movements were tracked as their arms were attached to an adjustable exoskeleton that did not hinder movement. In addition to intraindividual variability, other measures were also taken, such as mean distance between hand and target, reaction time across trials, etc. Correlation analyses revealed high correlation between different measures within the same task, but a marked lack of correlation between measures across tasks. However, across distinct experimental sessions, participants' values were similar, aligning with the observations of this dissertation that the ten participants who carried out both Production A and Production B produced speech with highly similar acoustic ranges and

rates of intraindividual variability. Kasuga et al. (2022) interpreted these results as evidence that an individual has distinct abilities across motor tasks, thus discarding the possibility of a subject-level, task-independent parameter of motor ability. The idea that variability is unique to the task itself and not a generalizable characteristic of an individual nor of the muscles involved was also observed by Streamer & Queen (2024), in a study in which they measured peak impact force and average loading rate (the speed at which force is applied) by means of in-shoe sensors during four tasks: level walking, stair ascent, stair descent and sit-to-stand motion. They observed distinct rates of variability for each of the tasks measured.

A more nuanced conclusion was reported by Oomen et al. (2023) who measured the Standard Deviation of motor variability across three tasks: carrying, lifting and simulated sawing. They found no strong correlation between variability of simulated sawing and lifting or simulated sawing and carrying. However, the presence of a moderate correlation between variability in carrying and lifting led them to conclude that variability may extend across tasks but only those that share the same degrees of freedom, i.e., the amount of possible movements or muscle combinations available to an individual for a specific task. In their study, the lifting and carrying tasks involved movement of the entire body and thus greater degrees of freedom, whereas the simulated sawing task only required use of the upper extremities, resulting in lower degrees of freedom. This would imply that finer motor tasks that offer less room for variability pattern differently than gross motor tasks. However, in addition to the shared degrees of freedom between the lifting and the carrying task, the former was essentially a shortened version of the latter, in that both involved lifting a crate from the floor (with the carrying task additionally requiring its displacement). Therefore, it is likely that both actions involved similar muscular movements in at least part of the task. This moderate correlation could be seen rather as a confirmation that similar levels of variability are observed for similar movement sequences, thus coinciding with the results from the previous studies reviewed here.

Additionally, rates of variability for a single task have been found to be consistent across experimental sessions. Sandlund et al. (2017) measured Standard Deviation of various parameters, such as elbow and wrist joint angles, thumb force and velocity, in a pipetting task that participants carried out on three separate days. Despite a large amount of interindividual

variability, intraindividual variability was relatively consistent across the three measurement sessions. Jackson et al. (2020) also observed consistent rates of motor variability across experimental sessions in participants performing a complex assembly task, agreeing with Kasuga et al. (2022)'s results. In sum, motor variability is largely independent of the task performed and does not represent an overall rate of variability for a single participant. Correlations are often observed, however, across variability in similar tasks that make use of the same limbs, implying that variability is dependent on specific movement configurations. These findings, coupled with the observed stable rates of variability in production across experimental sessions, suggest that each individual has a unique "motor signature", with unique motor abilities and rates of variability across different tasks.

The results of the current dissertation bolster the observations of the kinematic studies summarized here and offer additional convincing evidence of the task-independence and stability of intraindividual variability. Each speech articulatory configuration can be seen as a distinct task that involves the precise control of a specific set of articulators. The observation that similar rates of variability are only found in tasks that involve similar articulatory configurations or movements, coincides with similar rates of variability in Haar et al. (2017)'s study for reaching movements to distinct, yet spatially similar, targets and Oomen et al. (2023)'s study of similar variability in the lifting and carrying tasks, one of which being contained within the other. Together with the observations from neuroimaging studies, where similar brain regions are activated during speech, non-speech and general movement tasks, it is difficult to support the idea that speech articulation is fundamentally different from other motor tasks. The phonetic aspects of speech production are thus likely not part of a unique self-contained system, but are rather derived from, and likely mutually influencing, preexisting motor structures.

5.2 Individual variability in perception

As expected, participants differed considerably in terms of response consistency in the perception task. Intraindividual variability was measured across seven distinct continua: four stop voicing continua (/p-bɛ/, /p-bo/, /k-gɛ/, /k-go/), two vowel continua (/e-ɛ/ and /u-o/) and a

fricative continuum (/s-ʃ/). The Coefficient of Unalikeability (CoU) was not correlated across the four distinct continua (treating the four stop voicing continua as a single continuum), but the CoU of all possible pairs of the four individual stop continua, that differed by place of articulation and following vowel, were highly correlated (see Table 3.4 and Figure 3.10). These results can be understood as corresponding to the same structure as observed for production variability. Namely, intraindividual variability in perception is a consistent characteristic specific to certain acoustic properties corresponding to distinct articulatory configurations, rather than a speaker-specific generalizable trait. For stop voicing, the four continua tested participants' perception of the same acoustic properties (voicing proportion and post-release duration) and the precise values of both properties are the direct consequence of the timing of the abduction and adduction of the vocal folds. The strong correlation in variability across these four continua indicates that listeners viewed all four continua as corresponding to the same task and likely used the same categorization strategy throughout.

As a reminder, the vowel and fricative continua differed from the stop voicing continua in the type of acoustic information tested, namely durational for the stop voicing and spectral for the vowel and fricative continua. Given these differences, the lack of correlation in intraindividual variability between the stop continuum and the vowel and fricative continua is unsurprising. The lack of correlation in response consistency between the two vowel continua (/e-ɛ/ and /u-o/) may seem unexpected, given the similarity in task type. However, the large distance in acoustic values (largely F2) and thus tongue position upholding these two contrasts can explain the lack of correlation. Although in the production data, correlations were observed between the degree of variability of vowels with similar acoustic values in either the horizontal or vertical dimension (such as F1 of /o-e/), it is not possible to make such a comparison in perception data given that steps on the vowel continua differed in both F1 and F2 concurrently. Finally, the lack of correlation between the fricative continuum and either of the vowel continua is likely due to the large acoustic and articulatory differences between the two categories. In sum, the tendencies observed in the perception experiment provide further evidence that intraindividual variability is only similar across segments that share acoustic and articulatory commonalities.

In contrast to the results in this dissertation, various studies that have investigated response consistency in perception did find a significant correlation across phonetic contrasts (Myers et al. 2024; Kim et al. 2025). For example, Myers et al. (2024) used a VAS task and observed a moderate correlation between the consistency in identification of /b-d/, differing in F1 and F2 transitions, and identification of /s-ʃ/, differing in waveform shape. Kim et al. (2025) observed similar results on a VAS task where categorization consistency in English speakers was correlated across various continua for stop voicing, vowel identity and fricative identity. On the other hand, neither Fuhrmeister & Myers (2021) nor Zaharchuk et al. (2026) found a correlation between response consistency on a stop place of articulation continuum and a fricative continuum, aligning with the observations of the current dissertation. While the difference in the task procedure and the methods used to analyze participant responses make it difficult to directly compare results, it is important to consider what additional factors could be at the root of these conflicting findings.

A commonly underestimated factor affecting results of perception studies is participants' attention and task engagement. Indeed, despite the instructions given by the researcher, it is likely that each participant approaches the task in a different way or takes the experiment "seriously" to a different degree. The presence of such *attentive* or *inattentive* participants and their effect on data reliability has been widely discussed in the context of behavioral and social science research. Responses of attentive and inattentive participants have been shown to differ considerably, resulting in a relevant binary division of participants (Berinsky et al. 2014; Roth & Yakobi 2024). The degree of attentiveness is often assessed by the accuracy of participants in passing "screener questions" which require careful reading of the task instructions to choose the correct response, although the reliability of such screeners has also been called into question due to their potential in biasing participant responses (Berinsky et al. 2016; Hauser et al. 2018; Roth & Yakobi 2024). For example, Berinsky et al. (2014) observed that attentive participants spend more time completing questions on a survey, which is similar to the interpretation of this dissertation that the correlation between greater mean reaction time and response consistency on the perception task is indicative of attentiveness (see Figure 3.17). Discarding such inattentive participants, with the goal of reducing noise, leads to a disproportionate exclusion of participants

with certain sociodemographic characteristics, engendering dangerous biases and inaccurately representing human population (Berinsky et al. 2014). Given these concerns, it is important to not overgeneralize the results from a cognitive task, such as the 2AFC perception experiment employed in this dissertation, without recognizing the potential confounding role that participant attentiveness or attitude may play. In the current study, the combined analysis of reaction time and the production-perception link appear to offer a reliable indicator of participant attentiveness. The interpretation of the strength of the production-perception link and its possible informativeness in terms of participant attentiveness will be discussed further in Section 5.3.

In addition to their differing rates of attentiveness, participants may simply approach the perception task in different ways. Logie et al. (1996), for example, observed that at the group level, both word length and phonological similarity significantly affected performance on a verbal short-term memory task, where sequences containing longer words and/or phonologically similar words were recalled with lower accuracy. However, at the individual level, 43% of participants were only affected by one of these factors. No significant differences conditioned by age, sex or education were found, suggesting this is a purely individual characteristic. Participants were subsequently asked what strategy they used to carry out the memorization task and a total of six distinct strategies were identified. It was posited that effects of word length and phonological similarity are only observable if some level of phonological encoding is accessed, and this was corroborated by the fact that a much stronger effect size was observed for participants who reported repeating the words in their mind to enhance recall. Performance of participants who reported resorting to semantic cues or memorizing only the first letter of each word, was found to be much less affected by either word length or phonological similarity. Morrison et al. (2016) also observed that self-reported strategies differed significantly not only across participants but also within an individual participant's performance across distinct working memory tasks (e.g. immediate serial recall or item recognition). Observed strategies were similar to those reported in Logie et al. (1996), namely silent repetition, semantics, visualization, grouping of words, etc. In other words, the task itself does not necessarily force participants to employ a specific strategy and individuals are not consistent in the strategies they employ across

tasks. These results have important implications when attempting to report general tendencies observed in cognitive tasks, as the underlying mechanisms participants employ to complete the task can vary greatly. The presence of such a wide diversity in approaches to the same task makes it clear why a strong production-perception link was not observed for all participants in this dissertation and offers convincing evidence that the presence of such a link is more a task-specific strategy rather than an automatic linguistic response. Overall, it is pertinent to consider the real effect individual participant factors, such as attentiveness, motivation, familiarity, cognitive ability, etc., may have on the results and exercise greater caution when drawing far-reaching conclusions (Logie 2018).

In addition to such behavioral differences, the effect of the specific instructions given to participants has also been shown to impact participant responses. Schertz (2025), for example, analyzed participant's degree of imitation in an implicit and explicit imitation task involving both VOT duration in stops and F2 value in vowels. No correlation was observed between the degree of imitation of stops and vowels in the implicit imitation task. However, an individual's degree of implicit imitation was correlated for different segments within the same class, namely for both /p/ and /t/ in the VOT imitation task and /æ/ and /u/ in the F2 imitation task. In explicit imitation, however, Schertz (2025) observed a strong correlation across all measures, implying that the intentionality with which a participant carries out a task can significantly influence their behavior. Similar results have been observed in motor variability studies, where the constraints of a task differentially affect the amount of observed variability. Urbán et al. (2019) had participants carry out a pointing task under three conditions: no target, target and reward, and target and no reward. Besides the expected result of greater variability in the task with no target, a significantly lower degree of variability was observed in the target and reward task, indicative of additional, conscious effort exerted by participants in order to achieve the goal of obtaining the reward. Considering the combined role of task instruction and natural differences in participant attentiveness, the conflicting results across different studies regarding the correlation in intraindividual variability across acoustic dimensions can be best understood as being the result of both the unique sample used in each study and participants' attitude or strategies employed while conducting the task.

Further results from studies in cognitive sciences coincide with the observation in this dissertation that intraindividual variability in perception is only correlated across segments that share similar acoustic properties and articulatory configurations. Allaire & Marsiske (2005) observed no significant correlation between intraindividual variability across three different cognitive tasks, assessing inductive reasoning, perceptual speed, and list memory, implying that distinct cognitive abilities lead to distinct rates of variability. The degree of variability was, however, generally correlated across experimental sessions, indicating a stable characteristic for each individual within a given task. Similarly, Judd et al. (2024) measured intraindividual variability across eleven distinct cognitive tasks corresponding to three types: mathematical ability, working memory and nonverbal reasoning. Weak correlations were observed between different tasks within the working memory category, arguably since such tasks encourage similar strategies and access similar cognitive constructs. However, no type-internal correlations were observed for mathematical nor non-verbal tasks and, crucially, no correlations were observed between tasks belonging to distinct task types. Aschenbrenner & Jackson (2026) carried out a similar study measuring intraindividual variability across a wide range of different tasks (working memory, episodic memory and processing speed), including both mathematical and linguistic objects. They observed moderate type-internal correlations across variability in episodic memory and processing speed tasks, but no type-internal correlation between variability in working memory tasks, contrary to Judd et al. (2024)'s result. Again, they observed nearly no significant correlations between tasks belonging to different task types. These results are strong evidence of task-specific constraints on variability and go against the idea that certain individuals are more variable across all tasks. Given that the different continua in the perception experiment in this dissertation were likely perceived by participants as distinct tasks, due to their testing of different acoustic measures, it is highly plausible that individuals employed distinct strategies for different continua. This would explain why those participants that exclusively selected /ʃ/ on the fricative continuum, for example, were not the same participants who exclusively selected /o/ on the /u-o/ vowel continuum.

The interdisciplinary results presented here coincide with the observations in the current dissertation for both the production and perception task and the various motricity studies

discussed earlier (Getchell et al. 2001; Kasuga et al. 2022; Streamer & Queen 2024, etc.). Overall, intraindividual variability is specific to segments sharing a common articulatory configuration and/or acoustic consequence in both production and perception. This is likely due to cross-dimensional cognitive characteristics of variability that affect all aspects of human behavior and action and implies that understanding the structure of intraindividual variability is essential to better understand the mental organization of phonological information.

5.3 Variability in the production-perception link

Two types of links between production and perception were investigated: the relation between acoustic measures and the relation between intraindividual variability. The strength of the production-perception link between acoustic cues revealed various significant tendencies that merit further interpretation. The stop continuum was designed to ensure covariation between post-release aspiration duration and closure voicing proportion, both of which were shown to be highly variable in the production of the Laurentian French stop voicing contrast (see Figure 2.5 and Figure 2.6). A significant, group-level correlation was observed between production and perception cue weights for both cues, indicating that the relative importance of each of the two acoustic properties was similar in both modalities (see Figure 4.4). This implies that participants who consistently fully voice or nearly fully voice phonologically voiced stops in production assign greater importance to the presence of voicing in perception. Likewise, frequent partial-voicers in production rely less on voicing proportion in perception when categorizing stops. While this group-level tendency is certainly interesting, upon closer inspection of individual production-perception differences, a non-negligeable amount of interindividual variability was observed, indicating that the strength of the production-perception link varies considerably across individuals (see Figure 4.7).

A similar degree of interindividual variability regarding the production-perception link was observed for both vowel continua, where some participants' results imply that they closely mapped the model speaker's vowel space onto their own and then carried out categorization, while others appear to have essentially ignored their own vowel space and possibly relied on

distributional properties based solely on the stimuli they heard (see Figure 4.14). For the fricative continuum, most participants had very large production-perception differences, largely due to the fact that the majority of participants only made use of one of the two possible responses in the task (see Figure 4.16). This highlights in itself the consistent nature of interindividual differences and is a clear example of distinct subject-specific performances on a given task. Overall, no correlations were observed between the strength of the production-perception link across continua, implying that the presence of such a link is affected both by the task type (the specific continuum) and the phonological category being tested.

The strength of the production-perception link was found to be moderately correlated with both mean reaction time and the degree of intraindividual variability in the perception task for the vowel and fricative continua (although no correlation was observed between the CoU and the production-perception difference for the /e-ɛ/ contrast). That is, participants who were less consistent and slower responders in the perception task presented less distance between mean production and perception acoustic values. No correlation was observed between mean reaction time and production-perception difference for the stop continuum, which is likely due to the different task nature (cue weights vs. categorical boundary) or the different type of acoustic cues (durational for the stop voicing continua and spectral for the vowel and fricative continua), as discussed in Section 4.4. Participants with greater variability and longer reaction times likely compared each stimulus with their own production categories before making their decision. On the other hand, quicker, more consistent participants took less time in choosing a response based on their own categories and, instead, opted for an efficiency strategy where they either chose in a semi-random fashion or discerned the categorical boundary exclusively based on the model speakers' own repetitions. It is therefore possible to postulate that such a degree of production-perception difference may be an indirect way of measuring participant engagement with the task. At the very least, a production-perception difference measure may be able to determine the use of a particular strategy, i.e. actively referencing production during perception, with which an individual approached a 2AFC categorization task.

As outlined in Section 1.2.2, studies investigating the link between acoustic cues in production and perception have been inconclusive. While many studies find what seems like a

robust link, many others observe no significant connection. The results of this dissertation can begin to offer some clarity in the matter. The ability to observe a production-perception link between categorical boundaries or cue weights, while is likely partially dependent on the specific acoustic cue(s) investigated, is mostly contingent on a participant's approach to the task. It is therefore conceivable that even if some participants actively *choose* to make use of production categories while others do not, both groups can arrive at similar response patterns.

Regarding a possible link between intraindividual variability in production and perception, no group-level correlation was observed for any of the four perception continua (stop voicing, two vowel pairs and fricatives). Given the task-specificity of variability rates observed in this dissertation and the corresponding results in motor and cognitive studies, as presented in Section 5.1 and Section 5.2, this lack of correlation is relatively unsurprising. Producing speech and perceiving speech are two entirely different tasks that, while in some way undoubtedly access a common linguistic system, rely on largely separate physiological structures (vocal tract and hearing apparatus) and neurological processes (muscle control and interpretation of acoustic information). A similar lack of correlation between production and perception variability was observed by Cheng et al. (2021). Chao et al. (2019), however, observed a strong correlation between production and perception variability, although participants in their study heard manipulated tokens of their own voice, which arguably corresponds to a different task than that of perceiving the speech of others. Similarly, Franken et al. (2017) and Harper (2021) observed some degree of correlation between production and perception intraindividual variability. However, in both studies, perception variability was measured by an 4IAX discrimination task, where participants had to determine whether the second or third sound was different from the rest. Identification and discrimination are, again, two very different tasks that likely access different aspects of speech perception, making it unwise to directly compare them with the results of the current study. In addition, the results of both of these studies were only marginally significant and not present across all contrasts, suggesting a possible task- or segment-specific effect. Overall, most evidence points to a lack of a direct link between variability in production and perception. In sum, the relative independence of intraindividual variability across production and perception and the seemingly optional nature of a strong production-perception link for

acoustic cues both imply that production and perception are separate aspects of the linguistic system that *can* be, but are not automatically, linked.

5.4 Implications for language models

While the results of the studies in this dissertation only provide indirect evidence of neurological and motor processes, it remains pertinent to discuss their implications for models of speech production and perception. As discussed in Section 1.2.1, a plethora of models explaining possible connections between production and perception have been proposed, tested, redesigned and retested over the past half a century, such as Motor Theory (Liberman & Mattingly 1985), Direct Realism (Fowler 1986), WEAVER++ (Levelt 1993), exemplar theories (Nearey 1997; Pierrehumbert et al. 2002), the PRESENCE model (Moore 2007), and many more. Although all of these models have enjoyed some empirical support, none is free of theoretical drawbacks, and experimental observations that challenge a model's precepts have been reported for all existing models. The results of the current dissertation clearly set two important conditions for any model aiming to describe the relationship between production and perception. First, any direct connection between the two modalities is not automatic, given the high degree of interindividual variability in the strength of the production-perception link and considerable intraindividual variability across phonological categories described in Chapter 4. Second, intraindividual variability is not only a ubiquitous occurrence but, more importantly, is not random but rather is a structured and stable property of an individual. Given these two constraints, the type of model that is best able to explain the observations of the current studies is one that does not posit strict, discrete phonological units but rather some rendition of exemplars.

Specifically, a cognitive system composed of exemplars naturally does not divide phonological segments into entirely separate categories but permits a certain degree of fuzziness or overlap between groupings. The major drawback of exemplar theories is the unrealistic cognitive processing power required to store, access and carry out probability calculations on the mental databases containing every instance of produced and perceived speech (Baayen et al.

2013). Therefore, it is much more likely that through language use and exposure, distinct segments are somehow assigned a permitted amount of variability that form such 'exemplar clouds', or spaces of acceptable utterances, without the need to constantly store new language objects. A similar idea was proposed by Harper (2021), who also observed structured and consistent intraindividual variability across distinct segments in both articulation and acoustics. Harper (2021)'s proposed a modified version of Dynamic Field Theory to explain that phonological units can be both discrete and correspond to a distributional target space.

The question of how intraindividual variability is incorporated into production and perception merits continued investigation. However, the observations of the current dissertation reveal that the extent of intraindividual variability for a given segment is specific to each speech dimension. In production, the degree of intraindividual variability is likely determined by subject-specific differences in articulatory precision. In the same way that individuals differ widely from one another in the degree of variability of everyday activities, such as walking (Streamer & Queen 2024), clapping (Getchell et al. 2001) or lifting objects (Oomen et al. 2023), and the degree of such motor variability is not correlated across distinct tasks, it is likely that in speech articulation as well, individuals are simply able to produce certain articulatory configurations with greater accuracy. The reasons behind this are likely a combination of frequency of use of certain articulations and semi-random genetic and/or physiological factors. These physiological factors do not necessarily equate to anatomical differences. Indeed, Harper (2021) observed few consistent correlations between differences in vocal tract morphology, such as pharynx length or palate shape, and the degree of acoustic or articulatory variability. Instead, it is more likely that these unique, individual patterns correspond more to differences in neural dynamics controlling the activation of distinct muscle groups. Preliminary evidence for such interindividual differences in articulatory control was observed by Tan et al. (2026) who used electromyography to identify the muscle synergies, i.e., functional groups of muscles that work together to complete a movement, used in speech articulation. Although certain similarities were observed between subjects, there remained a large degree of interindividual variability, implying unique roles of certain groups of muscles dependent on individual characteristics.

Intraindividual variability in perception is likely incorporated into the representation of speech segments in a similar way, due to both experiential factors, such as language exposure and attitude, and cognitive and physiological differences in the auditory system. In other words, some individuals simply have greater tolerance for ambiguity surrounding certain acoustic properties and/or phonological segments. Various studies have shown that some listeners are simply better at distinguishing linguistic sounds than others (Díaz et al. 2008; 2016). Similarly, it has been found that individuals with musical abilities are consistently better than non-musicians at various perception tasks, such as speech-in-noise perception and the detection of metric structure (Marie et al. 2011; Kraus et al. 2012; Başkent & Gaudrain 2016). However, it appears that such correlations may be due in part to underlying genetic differences where the supposed musician advantage in speech perception is not due to explicit musical training, per se, but rather the presence of a genetic predisposition to increased musical aptitude fosters an interest in music which in turn enhances preexisting perception abilities (Schellenberg 2015). Such results again underscore the highly individualized nature of precise perception tendencies which undoubtedly affect the exact architecture of mental representations of phonological and phonetic information.

While the precise cognitive organization of phonological information cannot be determined solely by acoustic and behavioral tasks, such as the production and perception experiments in this dissertation, the highly structured intraindividual variability observed here suggests that speech segments are likely organized according to articulatory configurations and possibly their resulting acoustic consequences, rather than abstract phonological categories. This conclusion is supported by Mugler et al. (2018) whose neuro-imaging data revealed more similar activation patterns for segments with similar articulatory gestures, rather than segments belonging to the same phonological category. The large degree of inter- and intraindividual variability in production observed in Production B and in similar studies (Whalen et al. 2018; Harper 2021, etc.), as well as results of imitation studies that prove that speakers are capable of shifting their production categories with relative ease (Pardo 2006; Nielsen 2011; Kim & Clayards 2019; Schertz 2025, etc.), strongly suggest that tolerance for intraindividual variability is likely an integral part of the phonological grammar.

The structure of the linguistic system is likely connected to and codetermined by larger underlying organizational systems involving non-linguistic movement and non-linguistic audition. As described in Section 5.1, the brain regions activated during speech production appear to encompass both speech-specific, as well as more generalized motor-related neural patterns (Bouchard et al. 2013). Similar results have been observed for speech perception, where both speech-specific and general auditory-related brain regions are activated when hearing speech (Joanisse & Gati 2003; Price et al. 2005; Norman-Haignere et al. 2015; Zoefel & Kösem 2024). Such results underscore the interconnectedness of distinct functions and provide strong evidence that both speech production and perception are built upon existing cognitive structures that are also employed in other biological processes, thus implying that the precise organization of speech information likely varies across individuals.

Finally, the observation of strong connections between production and perception for some individuals show that the two modalities can likely be linked in some way, although it cannot be ruled out that the link between these participants' production and perception values is accidental. However, as discussed in Section 4.4 and Section 5.4, the degree of this connection is likely under an individual's control to a certain extent and is dependent on the specific approach used in the given perception task, emphasizing that such a link is not automatic nor an inherently necessary aspect of speech. It is also plausible that some individuals simply have a stronger link between production and perception than others, likely due to both experiential and genetic factors. Constant language exposure naturally enhances neural processes used in speech production and perception, in the same way that musical or athletic training enhances certain perceptual or motor functions. As mentioned previously, musicians have been shown to have an increased ability in discriminating speech in noise (Başkent & Gaudrain 2016) and in novel lexical tone discrimination (Choi et al. 2024) when compared to nonmusicians. While this ability may be due to a natural predisposition, the causal effect of increased musical training on increased perception skills is undebated (Schellenberg 2015). Similarly, expert athletes show more localized activity in distinct brain areas when compared to novice athletes, indicating enhanced motor specificity and efficiency (Kim et al. 2014; Peng et al. 2025). All in all, although all humans must develop specialized neural dynamics to produce and perceive language, varying life experiences,

approaches and attitudes towards language as well as individual genetic differences likely affect the precise makeup of these modalities across individuals. Therefore, the strength of a link between production and perception, in addition to being both segment- and task-dependent, likely also reflects individual differences in the strategies developed during language acquisition and use. Models of speech production and perception would therefore benefit from adopting a wider approach and incorporating more neural and behavioral results from a range of human experience to better capture the likely interconnected nature of such cognitive processes.

5.5 Role of variability

If one assumes the goal of linguistic communication is to ensure accurate comprehension of the transmitted information, production and perception variability, that is not the result of coarticulatory or sociophonetic processes, is often seen as an undesirable characteristic of speech. Similarly, variability in the motor domain or cognitive functions is often seen as detrimental to the consistency and predictability of such actions, thus heightening the risk of physical injury or cognitive fragility (Oomen et al. 2023; Aschenbrenner & Jackson 2026). However, as more attention is paid to intraindividual variability, evidence for the benefits of variability is increasing and more and more studies show that variability can be both advantageous and even necessary to ensure successful task completion.

Li et al. (2001) investigated intraindividual fluctuations in sensorimotor skills, such as duration and speed in walking tasks, verbal memory and spatial memory in older adults. Their results indicated that greater variability in sensorimotor skills was correlated with weaker scores on both the sensorimotor and memory tasks, indicating a detrimental effect of variability. However, greater variability on the spatial memory task, where individuals had to identify the newest object placed blindly on a board in front of them, was observed in participants with a higher level of accuracy for this task, suggesting an advantageous effect of variability. These results imply that variability can have both an adaptive and maladaptive function. *Maladaptive* means that greater variability results in weaker performance and is thought to reflect an increase in neuromuscular noise. This interpretation is further corroborated by the fact that in Li et al.

(2001)'s study, sensorimotor variability increased with age, thus representing a natural result of cognitive decline. On the other hand, *adaptive* response fluctuations are proposed to reflect an exploration of possible behaviors and strategies employed particularly on more difficult tasks, allowing the individual to test out different paths until settling on the optimal one. No correlation was observed between age and variability on the spatial memory task, indicating that the use of more or less effective strategies reflected individual differences rather than a consequence of age (Li et al. 2001). In other words, variability can be the result of either an active use of distinct task engagement strategies or a degree of uncontrolled inconsistency (Siegler 1994; Allaire & Marsiske 2025).

Cañigueral et al. (2023) observed a similar distinction between the role of adaptive and maladaptive variability in 6-13-year-old children who completed distinct cognitive processing tasks. Specifically, on a more difficult "Stopping" task, where individuals had to exercise inhibitory control by suppressing impulsive responses, greater variability was correlated with greater accuracy and greater overall improvement across multiple weeks of training. On the other hand, greater variability was correlated with lower accuracy on a "Going" task, where participants only had to respond as quickly as possible to an on-screen stimulus. Again, this seems to imply that in tasks that require unpredictable responses or some degree of trial-to-trial contemplation, variability is beneficial and is indicative of more flexible cognitive processing and response diversification. However, in less cognitively demanding tasks, which require simpler and more consistent responses, variability is a sign of lower performance and inefficient task response. The 2AFC perception task in this study is an example of a cognitively demanding task, similar to the "Stopping" task in Cañigueral et al. (2023)'s study, since participants must choose the most accurate response upon hearing each stimulus and this response is most often different from that of the previous trial. Although such a task does not allow for an accuracy score or a direct performance evaluation, given that there is no a priori correct response, the observation that more variable listeners were shown to have a stronger production-perception link across acoustic cues, can be taken as an indication of greater adaptability to the task. In this regard, participants with greater variability possess more flexible perception, which likely confers an advantage in adapting to new sounds or altered acoustic signals.

In a similar study, Klepacz et al. (2025) analyzed dispersion, i.e. intraindividual variability across tasks, over a range of cognitive functioning tasks, such as episodic memory, inductive reasoning and verbal fluency, carried out at two distinct intervals, nine years apart. While dispersion is not directly comparable to trial-to-trial variability, as measured in this dissertation, it still provides a useful measure of variability at the level of the individual. A greater level of dispersion indicates, for example, that an individual performed very well on some tasks and very poorly on others. Their results indicated that higher levels of dispersion were associated with higher overall performance and lower age. That is, individuals who perform well on some tasks and weaker on others obtained a higher cognition score than individuals who perform at a consistent level across tasks. The negative correlation with age, where younger participants have greater rates of dispersion, supports Li et al. (2001)'s observation that greater variability is not necessarily a natural process of cognitive decline or indicative of cognitive noise. In sum, task-dependent, greater variability is generally synonymous with greater flexibility, which is advantageous in real-life situations where one is constantly confronted with unexpected stimuli and the need to adapt to change.

Intraindividual variability likely can confer individuals with distinct advantages in speech production as well. As outlined in Section 5.1, motor control in speech articulation is similar and likely utilizes the same neural structures as movement in general, including that of muscles not related to speech. Various studies have reported that higher levels of motor variability equate to greater motor learning abilities (Wu et al. 2014; Dhawale et al. 2017; Knelange & López-Moliner 2020; Matsuda & Abe 2023). Wu et al. (2014), for example, observed that greater motor variability was highly correlated with faster motor learning across two tasks. In the first task, participants were instructed to perform reaching movements based on a shape provided to them at the start of the experiment. After 250 trials to establish a baseline, participants completed an additional 500-1000 trials and a score was indicated after each movement which the participants believed to reflect the accuracy of their movement. In reality, the score was deceptive and reflected the accuracy of the arm path towards a distinct target that did not correspond with the target participants had been shown. Higher rates of variability at baseline were correlated with faster rates of learning, that is, faster rates of adjusting the trajectory of the reaching movement

to match the new path that was deduced from the accuracy scores. In the second task, participants carried out reaching movements in an environment that produced outside forces on the hand, thus perturbing movement. In this task, participants were not provided with any feedback but naturally learned to adjust their movements to account for the external force and increase accuracy. Again, greater variability at baseline was highly correlated with faster learning, i.e., variable participants required less trials to successfully adapt their movements to the altered environment. As with variability in cognitive function tasks, motor variability is interpreted as the exploration of distinct strategies to gather knowledge that subsequently enables the optimal acquisition of the movement to be learned (Wu et al. 2014).

Such studies that find a correlation between motor variability and motor learning tend to use novel learning tasks to probe this link. This advantage has been found in linguistic tasks as well. During L2 acquisition, for example, exposure to increased variability improves learners' production and perception of novel sounds. This has been observed both in conditions where listeners are exposed to a diverse range of talkers as well as to a single talker with a wide range of acoustic values (Kartushina & Martin 2019; Wang & Feng 2025). Regarding production of one's native language, although infants are indeed learning novel motor skills when acquiring a language, adult native speakers can be said to have essentially mastered such movements, rendering any advantage of variability in learning novel tasks largely irrelevant. However, motor learning does not exclusively refer to the acquisition of novel movements, but also the adaptability of motor skills to new constraints, as modeled in the second experiment in Wu et al. (2014)'s study. Indeed, the state of the external environment and the internal condition of the human body is constantly in flux and a myriad of factors can affect the baseline state of the articulators employed in speech production, ranging from transient fatigue, boredom or sickness to more permanent conditions such as physical or cognitive damage. Adjusting motor output is thus a constant requirement even when executing motor actions that are performed on a daily basis (Matsuda & Abe 2023). Given such results for motor variability, it is conceivable that the participants in the current dissertation who produced greater rates of intraindividual variability in speech production may have a greater ease in learning novel speech sounds or in adjusting their own speech to different environmental factors.

Evidence for such advantages can be found in studies on sensorimotor adaptation to speech perturbations. In such adaptation studies, participants produce certain vowels and are unable to hear their natural productions, but instead they hear artificially modified versions of their own vowels. Specifically, F1 and F2 are often altered so as to push the vowel away from the intended target. Nault & Munhall (2020) observed that individuals with greater baseline formant variability have larger compensatory effects when faced with such formant perturbations. In other words, they adjust their productions more to adapt to what they now perceive as their new vowel space. Trudeau-Fisette et al. (2024) also observed that greater production variability was correlated with more accurate formant adaptation, although only in the speech of children and not in that of adults. They argued that children who are still in the process of acquiring language can benefit from greater production variability in that they can adapt their productions with more ease to successfully learn adult-like production patterns, again following the same explanations as proposed for motor and cognitive variability (Li et al. 2001; Wu et al. 2014). It appears then that intraindividual variability in production can have real benefits in both language acquisition and adaptation. Being more willing to adapt one's own production could have a social advantage as well, given that phonetic convergence, where individuals alter their speech to match that of their interlocutor, often serves to reduce social distance and increase other's perception of an individual as friendly or likeable (Babel 2009; Schweitzer et al. 2017). This connection was reported by Lee et al. (2021) who observed that when two speakers were instructed to converse together, the speaker with greater baseline variability was most often the speaker who produced greater phonetic convergence towards the speech of the other, less variable speaker. Having more variable speech production, and therefore less rigid categories, thus facilitates flexible adaptation of one's speech segments to achieve distinct communicational goals.

Overall, variability appears to confer multiple benefits to individuals in both motor and cognitive tasks, which both present strong similarities with the type of production and perception studies carried out here. The production experiment in the current study was a motor task as the articulation of distinct speech sounds involves the precise coordination of a range of muscular movement. The perception task probed various cognitive abilities, such as decision-making and resolution of ambiguity, both examples of cognitively demanding tasks where variability is

thought to be adaptive. Given the consistent presence of intraindividual variability in production and perception shown in this dissertation, and particularly the confirmation that variability is not simply random noise, but rather is a structured and stable element within specific regions of articulatory and acoustic space, it is crucial for phonetic research to move beyond the concept of "ideal language users" and to grant equal importance to both general, community-level tendencies and individual patterns.

6. Conclusion

This dissertation analyzed the presence and distribution of intraindividual variability in production and perception across stop voicing, vowels and coronal fricatives in Laurentian French. The consideration of a wide range of phonological categories that differ from one another both in the type and configuration of the main articulator (vocal folds for stop voicing and distinct tongue positions for vowels and fricatives) and the nature of the corresponding perceptually relevant acoustic cues (duration of closure voicing and post-release aspiration for stops, lower-frequency spectral information for vowel formants and higher-frequency spectral information for fricatives) allowed for broad conclusions to be drawn, highlighting that intraindividual variability is not random noise, but is structured and consistent within the production and perception of speech of an individual.

Chapter 2 presented the results of an acoustic experiment in which 49 participants produced multiple repetitions of frame sentences containing various segments of interest. Considerable interindividual variability was observed, implying unique cue uses for each participant. Nearly half of phonologically voiced stops were realized with incomplete closure voicing and subsequent post-release aspiration, although the prevalence of such patterns differed widely across participants. Regarding the vowel pairs tested (/e-ɛ/ and /u-o/), many participants produced overlapping vowel distributions, identified by large degrees of intraindividual variability. The fricatives /s/ and /ʃ/ were generally produced as distinct categories, although participants varied considerably in the average distance between acoustic values of the two. The magnitude of intraindividual variability for all segments, measured by means of the Coefficient of Variation, also differed widely across individuals and segments. However, within the speech of a single individual, variability was consistently correlated across segments that shared an articulatory configuration and therefore similar resulting acoustic values. This includes, for example, voicing proportion across different places of articulation of voiced stops, due to the shared laryngeal configuration, F1 and F2 between vowels that are produced with similar tongue height and/or frontness positions, respectively, such as between F1 variability of /u/ and /y/, and f0 across all vowels, regardless of the identity of the preceding segment, given a common glottal pattern to control pitch. Additionally, the magnitude of intraindividual variability was found to

be stable across recording sessions carried out months apart. Overall, intraindividual variability in production was found to be structured across distinct articulatory configurations and their acoustic consequences, implying that the organization of speech segments is also best described as corresponding to articulatory and acoustic targets rather than abstract phonological categories.

In Chapter 3, the results of the two-alternative forced choice identification experiment revealed similar tendencies as in production regarding the organization of intraindividual variability. The perception experiment was carried out by the same 49 participants immediately following completion of the production experiment and tested two types of continua: cue weighting of acoustic cues to stop voicing (closure voicing proportion and post-release aspiration duration across the four continua /p-bɛ/, /p-bo/, /k-gɛ/, /k-go/) and categorical boundaries between two phonemes (the vowel contrasts /e-ɛ/ and /u-o/ and the fricative contrast /s-ʃ/). A large amount of interindividual variability was observed for all continua, although some overall group tendencies were observed. Most participants were more sensitive to a lack of closure voicing for bilabial stops compared to velar stops. Inversely, the presence of long durations of post-release aspiration were more accepted as corresponding to a phonologically voiced stop for velars compared to bilabials. The division of the two vowel continua followed similar trajectories for most participants, albeit with considerably different transition points. The responses to the fricative continuum were clearly divided into three distinct groups: those that nearly exclusively selected only one of the options (either /s/ or /ʃ/) and those who showed a clear distinction between the two continuum endpoints. The result for the fricative continuum, in particular, underlies the importance of considering individual differences in task response and is evidence of highly distinct strategies employed by different participant groupings.

Intraindividual variability, quantified by means of the Coefficient of Unalikeability and corresponding to response consistency across trials, also varied greatly across individuals. Aligning with the results from the production experiment, similar degrees of variability were only observed between continua assessing acoustic properties that are direct consequences of similar articulatory configurations. Specifically, correlations were observed between all four stop voicing continua, but no correlation was observed between any of the other continua. The lack of

correlation between the two vowel continua, for example, again emphasizes that the organization of intraindividual variability is specific to precise articulatory regions, such as front and back vowels, rather than corresponding to larger phonological categories, such as vowels or coronals. Longer reaction times were found to be correlated with greater response inconsistency, i.e. greater intraindividual variability, for the vowel and fricative continua. This was proposed to be an indication of task-attentiveness, where individuals with greater variability likely attended to the acoustic properties of the stimuli with greater care. This hypothesis was further bolstered by a similar correlation between greater reaction time and a stronger link between production and perception categories.

A similar degree of interindividual variability to that observed for each modality separately was observed in the link between production and perception cue weights and categorical boundaries for stop voicing and vowels/fricatives, respectively, as reported in Chapter 4. Moreover, the difference between production and perception categories was not correlated across segments, again implying that the relationship between production and perception is likely not uniform across all speech but is structured differently for each phonological contrast or articulatory configuration. The individuality of the strength of the production-perception link could be interpreted in two ways. Either it corresponds to unique relationships for each individual between the two modalities or it is a reflection of each participant's specific approach to the perception task. In other words, some participants may have chosen to actively make use of their production cues to carry out the perception task, while others either opted for a different strategy, such as disambiguating stimuli solely based on the distributional properties of the acoustic cues of the model speaker, or simply had a greater level of distraction or task-inattentiveness. The strong correlations observed, again solely for the vowel and fricative continua, between longer reaction times and a stronger production-perception link as well as between greater variability and a stronger production-perception link offer support to the proposal that participants who make active use of their own production categories can be identified in that they require more time to reach a decision on this type of perception task and exhibit greater response inconsistency. This is likely due to their real-time processing of each stimulus even across identical trials.

Overall, while the degree of intraindividual variability in both production and perception differs greatly across language users, it is of a similar magnitude for segments sharing a common articulatory configuration and resulting acoustic values within the speech of an individual and is stable across time. This implies that intraindividual variability, possibly codified as a tolerance for flexibility, is a crucial part of the phonological system. Among the existing models on production and perception, an exemplar model is best suited at incorporating the results of the current dissertation. However, rather than every produced and perceived utterance being stored in the brain, the definition of the boundaries of a segment, determined both by articulatory configurations and the resulting acoustic consequences, are likely an integral part of such mental representations.

Studies investigating intraindividual variability in motor skills or cognitive tasks have obtained highly similar results to those of this dissertation, concluding that the magnitude of intraindividual variability is dependent both on the individual and the specific task carried out. Crucially, the observation of the ubiquitous nature of intraindividual variability across both linguistic and non-linguistic domains suggests that such variability across tasks makes up a unique and individual-identifying characteristic. In other words, intraindividual variability is not a random occurrence but a structured trait offering specific advantages to an individual for crucial skills such as learning and adaptability. These results suggest that speech production and perception, at least at the phonetic level, likely correspond to functions derived from preexisting cognitive structures and physical capacities. Moving forward, it will be beneficial to continue studying language from a cross-disciplinary perspective to better understand how it may interact with other aspects of motricity, cognition and additional aspects of human biology. While doing so, both community-level and individual patterns, as well as deviations from these patterns expressed as intraindividual variability, are essential to its understanding.

References

Ahn, S. 2018a. The role of tongue position in laryngeal contrasts: An ultrasound study of English and Brazilian Portuguese. *Journal of Phonetics 71*, 451-467.

Ahn, S. 2018b. *The Role of Tongue Position in Voicing Contrasts in Cross-Linguistic Contexts*. Ph.D. dissertation, New York University.

Al-Tamimi, J., & Khattab, G. 2018. Acoustic correlates of the voicing contrast in Lebanese Arabic singleton and geminate stops. *Journal of Phonetics 71*, 306-325.

Allaire, J. C., & Marsiske, M. 2005. Intraindividual variability may not always indicate vulnerability in elders' cognitive performance. *Psychology and aging 20*(3), 390.

Apfelbaum, K. S., Kutlu, E., McMurray, B., & Kapnoula, E. C. 2022. Don't force it! Gradient speech categorization calls for continuous categorization tasks. *The Journal of the Acoustical Society of America 152*(6), 3728-3745.

Arjmandi, M. K., & Behroozmand, R. 2024. On the interplay between speech perception and production: insights from research and theories. *Frontiers in Neuroscience 18*, 1347614.

Articulate Instruments Ltd. 2012. Articulate Assistant Advanced User Guide: Version 2.14. Edinburgh, UK: Articulate Instruments Ltd.

Aschenbrenner, A. J., & Jackson, J. J. 2026. The construct validity of daily cognitive variability. *Psychology and aging 41*(2), 205-216.

Audibert, N., & Fougeron, C. 2022. Intra-speaker phonetic variation in read speech: comparison with inter-speaker variability in a controlled population. *Proc. Interspeech 2022*, 4755-4759.

Auger, J. 2005. Un bastion francophone en Amérique du Nord: le Québec. In Valdman, A., Auger, J. & Piston-Hatlen, D. (eds.) *Le français en Amérique du Nord: état présent*, 39-80.

Baayen, R. H., Hendrix, P., & Ramscar, M. 2013. Sidestepping the combinatorial explosion: An explanation of n-gram frequency effects based on naive discriminative learning. *Language and speech 56*(3), 329-347.

Babel, M. E. 2009. *Phonetic and social selectivity in speech accommodation*. Ph.D. dissertation, University of California, Berkeley.

Babel, M. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics 40*(1), 177-189.

Babel, M., & Munson, B. 2014. Producing Socially Meaningful Linguistic Variation. In Goldrick, M., Ferreira, V. S. & Miozzo, M. (eds.), *The Oxford Handbook of Language Production*, Oxford Library of Psychology, 308-326.

Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., ... & Altenmüller, E. 2006. Shared networks for auditory and motor processing in professional pianists: evidence from fMRI conjunction. *Neuroimage 30*(3), 917-926.

Basilakos, A., Smith, K. G., Fillmore, P., Fridriksson, J., & Fedorenko, E. 2018. Functional characterization of the human speech articulation network. *Cerebral Cortex 28*(5), 1816-1830.

Bakst, S. 2021. Palate shape influence depends on the segment: Articulatory and acoustic variability in American English /ɹ/ and /s/. *The Journal of the Acoustical Society of America 149*(2), 960-971.

Başkent, D., & Gaudrain, E. 2016. Musician advantage for speech-on-speech perception. *The Journal of the Acoustical Society of America 139*(3), EL51-EL56.

Beddor, P. S. 2009. A coarticulatory path to sound change. *Language 85*(4), 785-821.

Beddor, P. S., Coetzee, A. W., Styler, W., McGowan, K. B., & Boland, J. E. 2018. The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. Language 94(4), 931-968.

Berinsky, A. J., Margolis, M. F., & Sances, M. W. 2014. Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. *American journal of political science 58*(3), 739-753.

Berinsky, A. J., Margolis, M. F., & Sances, M. W. 2016. Can we turn shirkers into workers?. *Journal of Experimental Social Psychology 66*, 20-28.

Blackwood Ximenes, A., Shaw, J. A., & Carignan, C. 2017. A comparison of acoustic and articulatory methods for analyzing vowel differences across dialects: Data from American and Australian English. *The Journal of the Acoustical Society of America 142*(1), 363-377.

Blanco, C. P., Bannard, C., & Smiljanic, R. 2016. Differences in the association between segment and language: Early bilinguals pattern with monolinguals and are less accurate than late bilinguals. *Frontiers in Psychology* 7, 993.

Boersma, P. & Weenink, D. 2022. Praat: doing phonetics by computer [computer program]. Version 6.3.02. Retrieved from: https://www.praat.org/

Bohland, J. W., & Guenther, F. H. 2006. An fMRI investigation of syllable sequence production. *Neuroimage 32*(2), 821-841.

Bonini, L., Rotunno, C., Arcuri, E., & Gallese, V. 2022. Mirror neurons 30 years later: implications and applications. *Trends in cognitive sciences 26*(9), 767-781.

Bouchard, K. E., Mesgarani, N., Johnson, K., & Chang, E. F. 2013. Functional organization of human sensorimotor cortex for speech articulation. *Nature 495*(7441), 327-332.

Brajot, F. X., Mollaei, F., Callahan, M., Klein, D., Baum, S. R., & Gracco, V. L. 2013. Articulatory phonetics of coronal stops in monolingual and simultaneous bilingual speakers of Canadian French and English. *Proceedings of Meetings on Acoustics 19*(1), 060064.

Brunelle, M., & Jannedy, S. 2007. Social effects on the perception of Vietnamese tones. *Proceedings of the 16th international congress of phonetic sciences*, 1461-1464.

Brunelle, M., Tấn, T. T., Kirby, J., & Giang, Đ. L. 2020. Transphonologization of voicing in Chru: Studies in production and perception. *Laboratory Phonology*, *11*(1), 1-33.

Brunelle, M., Brown, J., & Thị Thu Hà, P. 2022. Northern Raglai voicing and its relation to Southern Raglai register: Evidence for early stages of registrogenesis. *Phonetica 79*(2), 151-188.

Brunelle, M., Leb, K., Tạ, T. T., & Đinh, L. G. 2024. Voicing or register in Jarai dialects? Implications for the reconstruction of Proto-Chamic and for registrogenesis. *Journal of the International Phonetic Association 54*(2), 635-676.

Brunelle, M., Đinh, L. G., & Tạ, T. T. 2025. Redundant voicing and register in Mnong Râlâm. *Phonetica 82*(1), 45-85.

Brunelle, M., Schweizer, D., & Ahn, S. 2025. An Ultrasound Study of the Role of Larynx Height and Tongue Movement in the Production of Javanese Registers (Stiff-Slack Stops). *Journal of the Southeast Asian Linguistics Society JSEALS 18.2*, 93-119.

Brunner, J., Fuchs, S., & Perrier, P. 2009. On the relationship between palate shape and articulatory behavior. *The Journal of the Acoustical Society of America 125*(6), 3936-3949.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., ... & Freund, H. J. 2001. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European journal of neuroscience 13*(2), 400-404.

Bunker, L. D., & Hillis, A. E. 2022. Vascular syndromes: Revisiting classification of poststroke aphasia. In Argye, E. H. & Fridriksson, J. (eds.) *Handbook of clinical neurology 185: Aphasia*, 37-55.

Cañigueral, R., Ganesan, K., Smid, C. R., Thompson, A., Dosenbach, N. U., & Steinbeis, N. 2023. Intra-individual variability adaptively increases following inhibition training during middle childhood. *Cognition 239*, 105548.

Caporaso, J. G., Lauber, C. L., Costello, E. K., Berg-Lyons, D., Gonzalez, A., Stombaugh, J., ... & Knight, R. 2011. Moving pictures of the human microbiome. *Genome biology 12*(5), R50.

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B. & Carbone, E. 1973. The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *The Journal of the Acoustical Society of America 54*(2), 421-428.

Caramazza, A. & Yeni-Komshian, G. H. 1974) Voice onset time in two French dialects. *Journal of Phonetics 2*(3), 239-245.

Casserly, E. D., & Pisoni, D. B. 2010. Speech perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science 1*(5), 629-647.

CFPQ (Corpus de français parlé au Québec). CATIFQ, Université de Sherbrooke. Retrieved from: https://applis.flsh.usherbrooke.ca/cfpq/index.php/site/index

Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., Ford, C., Volcic, R. & De Rosario, H., 2017. *pwr: Basic functions for power analysis*.

Chang, S. E., Kenney, M. K., Loucks, T. M., Poletto, C. J., & Ludlow, C. L. 2009. Common neural substrates support speech and non-speech vocal tract gestures. *Neuroimage 47*(1), 314-325.

Chantefort, P. 1976. Diglossie au Québec limites et tendances actuelles. *Langue française* (31), 91-104.

Chao, S. C., Ochoa, D., & Daliri, A. 2019. Production variability and categorical perception of vowels are strongly linked. *Frontiers in Human Neuroscience 13*, 96.

Cheng, H. S., Niziolek, C. A., Buchwald, A., & McAllister, T. 2021. Examining the relationship between speech perception, production distinctness, and production variability. *Frontiers in Human Neuroscience 15*, 660948.

Cheung, C., Hamilton, L. S., Johnson, K., & Chang, E. F. 2016. The auditory representation of speech sounds in human motor cortex. *elife 5*, e12577.

Cho, T., & Ladefoged, P. 1999. Variation and universals in VOT: evidence from 18 languages. *Journal of phonetics 27*(2), 207-229.

Chodroff, E., & Wilson, C. 2017. Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics 61*, 30-47.

Chodroff, E., & Wilson, C. 2020. Acoustic–phonetic and auditory mechanisms of adaptation in the perception of sibilant fricatives. *Attention, Perception, & Psychophysics 82*(4), 2027-2048.

Choi, W., Ling, C. L. K., & Wu, C. H. J. 2024. Musical advantage in lexical tone perception hinges on musical instrument: A comparison between pitched musicians, unpitched musicians, and nonmusicians. *Music Perception: An Interdisciplinary Journal*,*41*(5), 360-377.

Clayards, M. 2018. Differences in cue weights for speech perception are correlated for individuals within and across contrasts. *The Journal of the Acoustical Society of America 144*(3), EL172-EL177.

Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., & Wissing, D. 2018. Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics 66*, 185-216.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*. New York: Routledge.

Cooke, M., García Lecumberri, M. L., Barker, J., & Marxer, R. 2019. Lexical frequency effects in English and Spanish word misperceptions. *The Journal of the Acoustical Society of America 145*(2), EL136-EL141.

Côté, M. H. 2012. Laurentian French (Quebec): Extra vowels, missing schwas and surprising liaison consonants. In Gess, R., Lyche C. & Meisenburg, T. (eds.), *Phonological Variation in French*, John Benjamins Publishing Company, 235-274.

Daniloff, R. G., & Hammarberg, R. E. 1973. On defining coarticulation. *Journal of Phonetics 1*(3), 239-248.

Davidson, L. 2016. Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics* 54, 35-50.

De Ribaupierre, A., & Lecerf, T. 2018. On the importance of intraindividual variability in cognitive development. *Journal of Intelligence 6*(2), 17.

Dhawale, A. K., Smith, M. A., & Ölveczky, B. P. 2017. The role of variability in motor learning. *Annual review of neuroscience 40*, 479-498.

Díaz, B., Baus, C., Escera, C., Costa, A., & Sebastián-Gallés, N. 2008. Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proceedings of the National Academy of Sciences 105*(42), 16083-16088.

Díaz, B., Mitterer, H., Broersma, M., Escera, C., & Sebastian-Galles, N. 2016. Variability in L2 phonemic learning originates from speech-specific capabilities: An MMN study on late bilinguals. *Bilingualism: Language and Cognitio*n *19*(5), 955-970.

DiCanio, C. 2013. Spectral moments of fricative spectra [Praat script]. Retrieved from: https://www.acsu.buffalo.edu/~cdicanio/scripts.html

Diehl, R. L., Lotto, A. J., & Holt, L. L. 2004. Speech perception. *Annual Review of Psychology 55*(1), 149-179.

Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. 2015. Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics 49*, 77-95.

Dostie, G. 2016. Le Corpus de français parlé au Québec (CFPQ) et la langue des conversations familières: Exemple de mise à profit des données à partir d'un examen lexico-sémantique de la séquence je sais pas. *Corpus* 15.

Drager, K. 2011. Speaker age and vowel perception. *Language and Speech 54*(1), 99-121.

Dumas, D. & Boulanger, A. 1982. Les matériaux d'origine des voyelles fermées du français québécois. *Revue québécoise de linguistique 11*(2), 49-72.

Dumas, D. 1987. *Nos façons de parler : Les prononciations en français québécois.* Sillery, QC: Presses de l'Université du Québec.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. 1971. Speech perception in infants. *Science 171*(3968), 303-306.

Everitt, B. 1998. *The Cambridge Dictionary of Statistics.* Cambridge, UK: Cambridge University Press.

Flores, G. E., Caporaso, J. G., Henley, J. B., Rideout, J. R., Domogala, D., Chase, J., ... & Fierer, N. 2014. Temporal variability is a personalized feature of the human microbiome. *Genome biology 15*(12), 531.

Fougeron, C. 1999. Prosodically conditioned articulatory variations: A review. *U.C.L.A. Working Papers in Phonetics 97*, 1-74.

Fougeron, C. 2001. Articulatory properties of initial segments in several prosodic constituents in French. *Journal of phonetics 29*(2), 109-135.

Fowler, C. A. 1986. An event approach to the study of speech perception from a direct–realist perspective. *Journal of phonetics 14*(1), 3-28.

Fowler, C. A. 1996. Listeners do hear sounds, not tongues. *The Journal of the Acoustical Society of America 99*(3), 1730-1741.

Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. 2003. Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of memory and language 49*(3), 396-413.

Fowler, C. A., Sramko, V., Ostry, D. J., Rowland, S. A., & Hallé, P. 2008. Cross language phonetic influences on the speech of French–English bilinguals. *Journal of Phonetics 36*(4), 649-663.

Franken, M. K., Acheson, D. J., McQueen, J. M., Eisner, F., & Hagoort, P. 2017. Individual variability as a window on production-perception interactions in speech motor control. *The Journal of the Acoustical Society of America 142*(4), 2007-2018.

Freeman, V. 2023. Production and perception of prevelar merger: Two-dimensional comparisons using Pillai scores and confusion matrices. *Journal of phonetics 97*, 101213.

Friesner, M. 2010. Une prononciation «tsipéquement» québécoise?: La diffusion de deux aspects stéréotypés du français canadien. *The Canadian Journal of Linguistics / La revue canadienne de linguistique 55*(1), 27-53.

Fuhrmeister, P., & Myers, E. B. 2021. Structural neural correlates of individual differences in categorical perception. *Brain and language 215*, 104919.

Galantucci, B., Fowler, C. A., & Turvey, M. T. 2006. The motor theory of speech perception reviewed. *Psychonomic bulletin & review 13*(3), 361-377.

Gay, T. 1968. Effect of speaking rate on diphthong formant movements. *The Journal of the Acoustical Society of America 44*(6), 1570-1573.

Georgiou, G. P., & Giannakou, A. 2024. Acoustic Characteristics of Greek Vowels Produced by Adult Heritage Speakers of Albanian*. Acoustics 6,* 257-271.

Getchell, N., Forrester, L., & Whitall, J. 2001. Individual differences and similarities in the stability, timing consistency, and natural frequency of rhythmic coordinated actions. *Research Quarterly for Exercise and Sport 72*(1), 13-21.

Ghosh, S. S., Matthies, M. L., Maas, E., Hanson, A., Tiede, M., Ménard, L., ... & Perkell, J. S. 2010. An investigation of the relation between sibilant production and somatosensory and auditory acuity. *The Journal of the Acoustical Society of America 128*(5), 3079-3087.

Giovannone, N., & Theodore, R. M. 2021. Individual differences in the use of acoustic-phonetic versus lexical cues for speech perception. *Frontiers in Communication 6*, 691225.

Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological review 105*(2), 251.

Greenberg, J. H. 1956. The measurement of linguistic diversity. *Language 32*(1), 109-115.

Haar, S., Donchin, O., & Dinstein, I. 2017. Individual movement variability magnitudes are explained by cortical neural variability. *Journal of Neuroscience 37*(37), 9076-9085.

Harper, S. 2020. Consistency in difference: the relationship between articulatory variability and segment differentiation for individual speakers. Presented at Laboratory Phonology 17, Vancouver, Canada.

Harper, S., Goldstein, L., & Narayanan, S. 2020. Variability in individual constriction contributions to third formant values in American English /ɹ/. *The Journal of the Acoustical Society of America 147*(6), 3905-3916.

Harper, S. K. 2021. *Individual differences in phonetic variability and phonological representation*. Ph.D. dissertation, University of Southern California.

Hauser, D. J., Ellsworth, P. C., & Gonzalez, R. 2018. Are manipulation checks necessary?. *Frontiers in psychology 9*, 362650.

Hauser, I. 2021. Contrast implementation affects phonetic variability: A case study of Hindi and English stops. *Laboratory Phonology 12*(1).

Hauser, I. 2022. Speech sounds in larger inventories are not (necessarily) less variable. *The Journal of the Acoustical Society of America 152*(5), 2664-2674.

Hay, J., Warren, P., & Drager, K. 2006a. Factors influencing speech perception in the context of a merger-in-progress. *Journal of phonetics 34*(4), 458-484.

Hay, J., Nolan, A., & Drager, K. 2006b. From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review 23*(3), 351-379.

Hay, J., Drager, K., & Warren, P. 2010. Short-term exposure to one dialect affects processing of another. *Language and speech 53*(4), 447-471.

Hazan, V., & Rosen, S. 1991. Individual variability in the perception of cues to place contrasts in initial stops. *Perception & Psychophysics 49*(2), 187-200.

Heald, S. L., & Nusbaum, H. C. 2015. Variability in vowel production within and between days. *PLOS One 10*(9), e0136791.

Honda, C. T., Clayards, M., & Baum, S. R. 2024a. Exploring individual differences in native phonetic perception and their link to nonnative phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance 50*(4), 370.

Honda, C. T., Clayards, M., & Baum, S. R. 2024b. Individual differences in the consistency of neural and behavioural responses to speech sounds. *Brain Research*, *1845*, 149208.

Holt, L. L., & Lotto, A. J. 2014. The alluring but misleading analogy between mirror neurons and the motor theory of speech. *Behavioral & Brain Sciences 37*(2), 204-205.

Hui, B., & Jia, R. 2024. Reflecting on the use of response times to index linguistic knowledge in SLA. *Annual Review of Applied Linguistics*, 1-11.

Hullebus, M. A., Tobin, S., & Gafos, A. 2018. Speaker-specific structure in German voiceless stop voice onset times. In *Proceedings of Interspeech 2018*, 1403-1407.

Idemaru, K., Holt, L. L., & 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*(6), 3950-3964.

Iskarous, K., Shadle, C. H., & Proctor, M. I. 2011. Articulatory–acoustic kinematics: The production of American English /s/. *The Journal of the Acoustical Society of America 129*(2), 944-954.

Jackson, J. A., Srinivasan, D., & Mathiassen, S. E. 2020. Consistent individual motor variability traits demonstrated by females performing a long-cycle assembly task under conditions differing in temporal organisation. *Applied Ergonomics 85*, 103046.

Jacques, B. 1984. *Étude acoustique des consonnes du français parlé à Montréal: analyse spectrale, intensité, durée*. PhD dissertation, Université de Montréal.

Jacques, B. & Charbonneau, R. 1989. Les consonnes occlusives du français de Montréal; étude spectrographique de quelques paramètres acoustiques. In Delgado-Martins, M. R., Bothorel, A., Galdin, J.-C., Wioland F. & Zerting, J.-P. (eds.), *Mélanges de phonétique générale et expérimentale offerts à Pela Simon*. Strasbourg: Publications de l'Institut de phonétique de Strasbourg, 45-61.

Jacques, B. & Gurlekian, J. A. 1992. Étude comparée de quatre paramètres acoustiques des occlusives en espagnol de Buenos Aires et en français de Montréal. *Revue québécoise de linguistique 22*(1), 257-271.

Jo, I., Maule, H., & Brandt, A. J. 2024. Specific leaf area and its within-individual variation in understory evergreen and deciduous woody species in New Zealand. *Plant Ecology 225*(10), 1071-1078.

Joanisse, M. F., & Gati, J. S. 2003. Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. *Neuroimage 19*(1), 64-79.

Jongman, A., Wayland, R., & Wong, S. 2000. Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America 108*(3), 1252-1263.

Jongman, A., Wang, Y., & Kim, B. H. 2003. Contributions of semantic and facial information to perception of nonsibilant fricatives. *Journal of speech, language, and hearing research 46*(6), 1367-1377.

Judd, N., Aristodemou, M., Klingberg, T., & Kievit, R. 2024. Interindividual differences in cognitive variability are ubiquitous and distinct from mean performance in a battery of eleven tasks. *Journal of Cognition 7*(1), 45.

Kader, G. D., & Perry, M. 2007. Variability for categorical variables. *Journal of statistics education 15*(2).

Kapnoula, E. C., Winn, M. B., Kong, E. J., Edwards, J., & McMurray, B. 2017. Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance 43*(9), 1594.

Kartushina, N., & Martin, C. D. 2019. Talker and acoustic variability in learning to produce nonnative sounds: Evidence from articulatory training. *Language Learning 69*(1), 71-105.

Kasuga, S., Heming, E., Lowrey, C., & Scott, S. H. 2022. High intra-task and low inter-task correlations of motor skills in humans creates an individualized behavioural pattern. *Scientific Reports 12*(1), 20156.

Keating, P., Cho, T., Fougeron, C., & Hsu, C. S. 2003. Domain-initial strengthening in four languages. In Local, J., Ogden, R., & Temple, R. (eds.) *Laboratory phonology VI: Phonetic interpretation*. Cambridge: Cambridge University Press, 145-163.

Kendall, T., & Fridland, V. 2017. Regional relationships among the low vowels of US English: Evidence from production and perception. *Language Variation and Change 29*(2), 245-271.

Kier, W. M., & Smith, K. K. 1985. Tongues, tentacles and trunks: the biomechanics of movement in muscular-hydrostats. *Zoological journal of the Linnean Society*, *83*(4), 307-324.

Kim, D., & Clayards, M. 2019. Individual differences in the link between perception and production and the mechanisms of phonetic imitation. *Language, Cognition and Neuroscience 34*(6), 769-786.

Kim, H., McMurray, B., Sorensen, E., & Oleson, J. 2025. The consistency of categorization-consistency in speech perception. *Psychonomic bulletin & review*, 1-13.

Kim, W., Chang, Y., Kim, J., Seo, J., Ryu, K., Lee, E., ... & Janelle, C. M. 2014. An fMRI study of differences in brain activity among elite, expert, and novice archers at the moment of optimal aiming. *Cognitive and Behavioral Neurology*,*27*(4), 173-182.

Kirby, J. 2018. Praatsauce: Praat-based tools for spectral analysis [Praat script]. Version 0.2.4. Retrieved from: https://github.com/kirbyj/praatsauce.

Klepacz, L. M., Cerino, E. S., & Hamm, J. M. 2025. Characterizing cognitive dispersion and its correlates across the adult lifespan in MIDUS. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences 80*(6), gbae201.

Knelange, E. B., & López-Moliner, J. 2020. Increased error-correction leads to both higher levels of variability and adaptation. *PLOS One 15*(2), e0227913.

Kochetov, A., Tabain, M., Sreedevi, N., & Beare, R. 2018. Manner and place differences in Kannada coronal consonants: Articulatory and acoustic results. *The Journal of the Acoustical Society of America 144*(6), 3221-3235.

Kogan, V. V. 2024. Compactness of Native Vowel Categories in Monolingual, Bilingual, and Multilingual Speakers: Is Category Compactness Affected by the Number of Languages Spoken?. *Languages 9*(7), 238.

Komeilipoor, N., Tiainen, M., Tiippana, K., Vainio, M., & Vainio, L. 2016. Excitability of hand motor areas during articulation of syllables. *Neuroscience Letters 620*, 154-158.

Kong, E. J., Beckman, M. E., & Edwards, J. 2011. Why are Korean tense stops acquired so early?: The role of acoustic properties. *Journal of phonetics 39*(2), 196-211.

Kong, E. J., & Edwards, J. 2016. Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics 59*, 40-57.

Koops, C., Gentry, E., & Pantos, A. 2008. The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas. *University of Pennsylvania Working Papers in Linguistics 14*(2), 12.

Kraus, N., Strait, D. L., & Parbery-Clark, A. 2012. Cognitive factors shape brain networks for auditory skills: spotlight on auditory working memory. *Annals of the New York Academy of Sciences 1252*(1), 100-107.

Kuznetsova, A., Brockhoff, P., & Christensen, R. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software 82*, 1-26.

Lai, W., Wade, L., & Tamminga, M. 2022. Individual differences in simultaneous perceptual compensation for coarticulatory and lexical cues. *Linguistics Vanguard 8*(1), 29-39.

Lancheros, M., Jouen, A. L., & Laganaro, M. 2020. Neural dynamics of speech and non-speech motor planning. *Brain and Language 203*, 104742.

Lara, A. F., & Pillot-Loiseau, C. 2023. Phonetic convergence and vowel overlap in simultaneous bilinguals, language learners, and monolinguals. *Radical: A journal of Phonology 3*, 238-279.

Leblanc, G. 2012. *Une étude acoustique des voyelles orales susceptibles d'être diphtonguées en français québécois*. M.A. thesis, Université Laval.

Lee, B., & Sidtis, D. V. L. 2017. The bilingual voice: Vocal characteristics when speaking two languages across speech tasks. *Speech, Language and Hearing 20*(3), 174-185.

Lee, S., Yildirim, S., Kazemzadeh, A., & Narayanan, S. S. 2005. An articulatory study of emotional speech production. *Interspeech*, 497-500.

Lee, Y., Goldstein, L., Parrell, B., & Byrd, D. 2021. Who converges? Variation reveals individual speaker adaptability. *Speech Communication 131*, 23-34.

Lenth, R. & Piaskowski, J. 2022. emmeans: Estimated marginal means. Aka Least-Squares Means [R package]. Version 1.8.2.

Leung, K. K., & Wang, Y. 2020. Production-perception relationship of Mandarin tones as revealed by critical perceptual cues. *The Journal of the Acoustical Society of America 147*(4), EL301-EL306.

Leung, K. K., & Wang, Y. 2024. Modelling Mandarin tone perception-production link through critical perceptual cues. *The Journal of the Acoustical Society of America 155*(2), 1451-1468.

Levelt, W. J. 1993. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J., Roelofs, A., & Meyer, A. S. 1999. A theory of lexical access in speech production. *Behavioral and brain sciences 22*(1), 1-38.

Levy, H., & Hanulíková, A. 2019. Variation in children's vowel production: Effects of language exposure and lexical frequency. *Laboratory Phonology 10*(1), 1-26.

Li, F., Munson, B., Edwards, J., Yoneyama, K., & Hall, K. 2011. Language specificity in the perception of voiceless sibilant fricatives in Japanese and English: Implications for cross-language differences in speech-sound development. *The Journal of the Acoustical Society of America 129*(2), 999-1011.

Li, P., Flege, J. E., Martin, C. D., & Kartushina, N. 2023. Speech Stability Across Time: Evidence from Norwegian Vowels in Spontaneous Speech Production. In *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS)*, 1177-1181.

Li, P., Martin, C. D., & Kartushina, N. 2026. Same vowels, distinct sounds: Phonetic differentiation in early Basque–Spanish bilinguals' vowel production. *Linguistic Approaches to Bilingualism*, 1-15.

Li, S. C., Aggen, S. H., Nesselroade, J. R., & Baltes, P. B. 2001. Short-term fluctuations in elderly people's sensorimotor functioning predict text and spatial memory performance: The MacArthur Successful Aging Studies. *Gerontology 47*(2), 100-116.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. 1967. Perception of the speech code. *Psychological review 74*(6), 431.

Liberman, A. M., & Mattingly, I. G. 1985. The motor theory of speech perception revised. *Cognition 21*(1), 1-36.

Lisker, L., & Abramson, A. S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word 20*(3), 384-422.

Liuzzi, G., Ellger, T., Flöel, A., Breitenstein, C., Jansen, A., & Knecht, S. 2008. Walking the talk—Speech activates the leg motor cortex. *Neuropsychologia 46*(11), 2824-2830.

Lo, R. Y. H., Liu, S., Vaughn, C., McAuliffe, M., & Babel, M. 2023. Variation in Perception and Production of /s/-/S/ in English. *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS)*.

Logie, R. H., Sala, S. D., Laiacona, M., Chalmers, P., & Wynn, V. 1996. Group aggregates and individual reliability: The case of verbal short-term memory. *Memory & Cognition 24*(3), 305-321.

Logie, R. 2018. Human cognition: Common principles and individual variation. *Journal of Applied Research in Memory and Cognition 7*(4), 471-486.

Lotto, A. J., Hickok, G. S., & Holt, L. L. 2009. Reflections on mirror neurons and speech perception. *Trends in cognitive sciences 13*(3), 110-114.

Louër, Dominique. 2019. *Le VOT dans une situation de contact des langues : Étude comparative de la production orale des locuteurs natifs et des apprenants du français*. M.S. thesis, University of Waterloo.

Lyons, E. M., Goldsworthy, M. R., & Hordacre, B. 2025. Exploring neural variability, movement variability and motor learning in healthy adults and people with stroke. *Clinical Neurophysiology 175*, 2110759.

MacKenzie, L., & Sankoff, G. 2010. A quantitative analysis of diphthongization in Montreal French. *U. Penn Working Papers in Linguistics 15*(2), 91-100.

MacLeod, A. A. N. 2006. *Production and perception of VOT and high vowels by bilingual and monolingual speakers of Canadian English and Canadian French*. Ph.D. dissertation, University of Washington.

MacLeod, A. A. & Stoel-Gammon, C. 2005. Are bilinguals different? What VOT tells us about simultaneous bilinguals. *Journal of Multilingual Communication Disorders 3*(2), 118-127.

MacLeod, A. A., & Stoel-Gammon, C. 2009. The use of voice onset time by early bilinguals to distinguish homorganic stops in Canadian English and Canadian French. *Applied Psycholinguistics*, *30*(1), 53-77.

Makashay, M. J. 2003. *Individual differences in speech and non-speech perception of frequency and duration*. Ph.D. dissertation, The Ohio State University.

Marie, C., Magne, C., & Besson, M. 2011. Musicians and the metric structure of words. *Journal of cognitive neuroscience 23*(2), 294-305.

Martin, P. 2002. Le système vocalique du français du Québec. De l'acoustique à la phonologie. *La linguistique* 38(2), 71-88.

Mathôt, S., Schreij, D., & Theeuwes, J. 2012. OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods 44*(2), 314-324.

Matsuda, N., & Abe, M. O. 2023. Error size shape relationships between motor variability and implicit motor adaptation. *Biology 12*(3), 404.

McAllister Byun, T., & Tiede, M. 2017. Perception-production relations in later development of American English rhotics. *PLOS One 12*(2), e0172022.

McGettigan, C., & Tremblay, P. 2018. Links between perception and production: examining the roles of motor and premotor cortices in understanding speech. *The Oxford handbook of psycholinguistics*, 306-332.

McGurk, H., & MacDonald, J. 1976. Hearing lips and seeing voices. *Nature 264*, 746-748.

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. 2007. The essential role of premotor cortex in speech perception. *Current Biology 17*(19), 1692-1696.

Meltzoff, A. N. 1993. The centrality of motor coordination and proprioception in social and cognitive development: From shared actions to shared minds. *Advances in Psychology 97*, 463-496.

Mohr, J. P., Pessin, M. S., Finkelstein, S., Funkenstein, H. H., Duncan, G. W., & Davis, K. R. 1978. Broca aphasia: pathologic and clinical. *Neurology 28*(4), 311-311.

Moineau, S., Dronkers, N. F., & Bates, E. 2005. Exploring the processing continuum of single-word comprehension in aphasia. *Journal of Speech, Language, and Hearing Research 48*(4), 884-896.

Molenberghs, P., Cunnington, R., & Mattingley, J. B. 2012. Brain regions with mirror properties: a meta-analysis of 125 human fMRI studies. *Neuroscience & biobehavioral reviews 36*(1), 341-349.

Montazeri-Najafabady, N. 2025. From One-Size-Fits-All to Precision Medicine: The Promise of Personalized Probiotics. *Probiotics and Antimicrobial Proteins*, 1-34.

Moore, R. K. 2007. Spoken language processing: piecing together the puzzle. *Speech communication 49*(5), 418-435.

Mora, J. C., Keidel, J. L., & Flege, J. E. 2015. Effects of Spanish use on the production of Catalan vowels by early Spanish-Catalan bilinguals. *The phonetics-phonology interface: Representations and methodologies*, 33-53.

Morrison, A. B., Rosenbaum, G. M., Fair, D., & Chein, J. M. 2016. Variation in strategy use across measures of verbal working memory. *Memory & cognition 44*(6), 922-936.

Mugler, E. M., Tate, M. C., Livescu, K., Templer, J. W., Goldrick, M. A., & Slutzky, M. W. 2018. Differential representation of articulatory gestures and phonemes in precentral and inferior frontal gyri. *Journal of Neuroscience 38*(46), 9803-9813.

Myers, E., Phillips, M., & Skoe, E. 2024. Individual differences in the perception of phonetic category structure predict speech-in-noise performance. *The Journal of the Acoustical Society of America 156*(3), 1707-1719.

Narayanan, S. S., Alwan, A. A., & Haker, K. 1995. An articulatory study of fricative consonants using magnetic resonance imaging. *The Journal of the Acoustical Society of America 98*(3), 1325-1347.

Nault, D. R., & Munhall, K. G. 2020. Individual variability in auditory feedback processing: Responses to real-time formant perturbations and their relation to perceptual acuity. *The Journal of the Acoustical Society of America 148*(6), 3709-3721.

Nearey, T. M. 1997. Speech perception as pattern recognition. *The Journal of the Acoustical Society of America 101*(6), 3241-3254.

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. 2004. Lexique 2: A new French lexical database. *Behavior Research Methods, Instruments, & Computers 36*(3), 516-524.

Newman, R. S., Clouse, S. A., & Burnham, J. L. 2001. The perceptual consequences of within-talker variability in fricative production. *The Journal of the Acoustical Society of America 109(3),* 1181-1196.

Niedzielski, N. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of language and social psychology 18*(1), 62-85.

Nielsen, K. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics 39*(2), 132-142.

Noiray, A., Iskarous, K., & Whalen, D. H. 2014. Variability in English vowels is comparable in articulation and acoustics. *Laboratory phonology 5*(2), 271.

Norman-Haignere, S., Kanwisher, N. G., & McDermott, J. H. 2015. Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron 88*(6), 1281-1296.

Ohala, J. J. 1983. The origin of sound patterns in vocal tract constraints. In MacNeilage, P. F. (ed.) *The Production of Speech*. New York: Springer New York, 189-216.

Ohala, J. J. 2011. Accommodation to the Aerodynamic Voicing Constraint and its Phonological Relevance. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)*, 64-67.

Ohala, J. J. & Solé, M.-J. 2010. Turbulence and Phonology. In Fuchs, S., Toda, M. & Zygis, M. (eds.) *Turbulent sounds. An interdisciplinary guide*. Berlin: Mouton de Gruyter, 37-97.

Oomen, N. M., Graham, R. B., & Fischer, S. L. 2023. Exploring the role of task on kinematic variability and assessing consistency in individual responses across repetitive manual tasks. *Ergonomics 66*(6), 749-761.

Ou, J., Law, S. P., & Fung, R. 2015. Relationship between individual differences in speech processing and cognitive functions. *Psychonomic Bulletin & Review 22*, 1725-1732.

Ou, J., & Law, S. P. 2017. Cognitive basis of individual differences in speech perception, production and representations: The role of domain general attentional switching. *Attention, Perception, & Psychophysics 79*(3), 945-963.

Ou, J., & Yu, A. C. 2022. Neural correlates of individual differences in speech categorisation: Evidence from subcortical, cortical, and behavioural measures. *Language, Cognition and Neuroscience 37*(3), 269-284.

Ou, J., Xiang, M., & Yu, A. C. 2023. Individual variability in subcortical neural encoding shapes phonetic cue weighting. *Scientific Reports 13*(1), 9991.

Pape D. & Jesus, L.M. 2011 Devoicing of phonologically voiced obstruents: Is European Portuguese different from other Romance languages?. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)*, 1566-1569.

Pape, D., & Jesus, L. M. 2014. Production and perception of velar stop (de) voicing in European

Portuguese and Italian. *EURASIP Journal on Audio, Speech, and Music Processing*, 1-10.

Pardo, J. S. 2006. On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America 119*(4), 2382-2393.

Park, Y., Perkell, J. S., Matthies, M. L., & Stepp, C. E. 2019. Categorization in the perception of breathy voice quality and its relation to voice production in healthy speakers. *Journal of Speech, Language, and Hearing Research 62*(10), 3655-3666.

Peng, L., Xu, L., Zhang, Z., Wang, Z., Zhong, X., Wang, L., ... & Shao, Y. 2025. Classifying athletes and non-athletes by differences in spontaneous brain activity: a machine learning and fMRI study. *Brain Structure and Function 230*(5), 74.

Perkell, J. S. 1969. *Physiology of speech production: Results and implications of a quantitative cineradiographic study*. Research monograph, MIT.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., & Zandipour, M. 2004. The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *The Journal of the Acoustical Society of America 116*(4), 2338-2344.

Perkell, J. S., Lane, H., Ghosh, S., Matthies, M. L., Tiede, M., Guenther, F. H., & Ménard, L. 2008. Mechanisms of vowel production: auditory goals and speaker acuity. *Proceedings of the Eighth international Seminar on speech production,* 29-32.

Pierrehumbert, J., Gussenhoven, C., & Warner, N. 2002. Word-specific phonetics. *Laboratory Phonology* 7(1), 101-140.

Pierrehumbert, J. 2003. Probabilistic phonology: discrimination and robustness. In: Bod, R., Hay, J., Jannedy, S. (eds.), *Probabilistic Linguistics*. The MIT Press, 177-228.

Pinget, A. F., Kager, R., & Van de Velde, H. 2020. Linking Variation in Perception and Production in Sound Change: Evidence from Dutch Obstruent Devoicing. *Language and speech 63*(3), 660-685.

Piot, L., Nazzi, T., & Boll-Avetisyan, N. 2024. Infants' sensitivity to phonotactic regularities related to perceptually low-salient fricatives: a cross-linguistic study. *Frontiers in psychology 15*, 1367240.

Porter Jr, R. J., & Lubker, J. F. 1980. Rapid reproduction of vowel-vowel sequences: Evidence for a fast and direct acoustic-motoric linkage in speech. *Journal of Speech, Language, and Hearing Research 23*(3), 593-602.

Preston, J. L., & Koenig, L. L. 2011. Phonetic variability in residual speech sound disorders: Exploration of subtypes. *Topics in Language Disorders 31*(2), 168-184.

Price, C., Thierry, G., & Griffiths, T. 2005. Speech-specific auditory processing: where is it?. *Trends in cognitive sciences 9*(6), 271-276.

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. 2006. Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences 103*(20), 7865-7870.

R Core Team. 2022. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Retrieved from: https://www.R-project.org/

Ray, B. K., & Gilbert, A. 2019. Language continuities and ruptures in Ottawa-Gatineau, Canada. In Brunn, S. & Kehrein, R. (eds.) *Handbook of the changing world language map*. Springer, 1-20.

Read, B., Wolters, L., & Berinsky, A. J. 2022. Racing the clock: Using response time as a proxy for attentiveness on self-administered surveys. *Political Analysis 30*(4), 550-569.

Recasens, D., & Espinosa, A. 2009. Dispersion and variability in Catalan five and six peripheral vowel systems. *Speech Communication 51*(3), 240-258.

Redford, M., & Baese-Berk, M. 2023. Acoustic theories of speech perception. In Aronoff, M. (ed.) *Oxford Research Encyclopedia of Linguistics*, New York: Oxford Academic.

Reed, G. F., Lynn, F., & Meade, B. D. 2002. Use of coefficient of variation in assessing variability of quantitative assays. *Clinical and Diagnostic Laboratory Immunology 9*(6), 1235-1239.

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. 2009. Changes of mind in decision-making. *Nature 461*(7261), 263-266.

Riordan, C. J. 1977. Control of vocal-tract length in speech. *The Journal of the Acoustical Society of America 62*(4), 998-1002.

Riverin-Coutlée, J., Roy, J. P., & Gubian, M. 2023. Using Mahalanobis distances to investigate second dialect acquisition: A study on Quebec French. *Language and Speech 66*(2), 291-321.

Rizzi, R., & Bidelman, G. M. 2024. Functional benefits of continuous vs. categorical listening strategies on the neural encoding and perception of noise-degraded speech. *Brain research 1844*, 149166.

Rizzolatti, G., & Arbib, M. A. 1998. Language within our grasp. *Trends in neurosciences 21*(5), 188-194.

Robb, M., Gilbert, H., & Lerman, J. 2005. Influence of gender and environmental setting on voice onset time. *Folia Phoniatrica et Logopaedica 57*(3), 125-133.

Roelofs, A. 2005. Spoken word planning, comprehending, and self-monitoring: Evaluation of WEAVER++. In Hartsuiker, R. J., Bastiaanse, R., Postma, A. & Wijnen, F. (eds.) *Phonological encoding and monitoring in normal and pathological speech*. Psychology Press, 42-63.

Roth, Y., & Yakobi, O. 2024. Attention! Do we really need attention checks?. *Journal of Behavioral Decision Making 37*(2), e2377.

Rudy, K., & Yunusova, Y. 2013. The effect of anatomic factors on tongue position variability during consonants. *Journal of Speech, Language, and Hearing Research 56*(1), 137-149.

Ryalls, J., Larouche, A., Dupont, C. & Fournier, A. 1990. Normes phonétiques de base pour l'enfant franco-québécois; Basic phonetic norms for the French Québec child. *Journal of Speech Language Pathology and Audiology/Revue d'Orthophonie et d'Audiologie 14*, 65-70.

Ryalls, J., Cliché, A., Fortier-Blanc, J., Coulombe, I. & Prud'hommeaux, A. 1997. Voice-onset time in younger and older French-speaking Canadians. *Clinical linguistics & phonetics 11*(3), 205-212.

Salinas Marchant, C. L. 2025. *Development of fricative speech sounds in French: A view through different acoustic and statistical analyses*. Ph.D. dissertation, University of Alberta.

Sandlund, J., Srinivasan, D., Heiden, M., & Mathiassen, S. E. 2017. Differences in motor variability among individuals performing a standardized short-cycle manual task. *Human movement science 51*, 17-26.

Schellenberg, E. G. 2015. Music training and speech perception: A gene–environment interaction. *Annals of the New York Academy of Sciences 1337*(1), 170-177.

Schertz, J. L. 2014. *The structure and plasticity of phonetic categories across languages and modalities*. Ph.D. dissertation, The University of Arizona.

Schertz, J., Cho, T., Lotto, A., & 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.

Schertz, J., & Clare, E. J. 2020. Phonetic cue weighting in perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science 11*(2), e1521.

Schertz, J., Adil, F., & Kravchuk, A. 2023. Underpinnings of explicit phonetic imitation: Perception, production, and variability. *Glossa Psycholinguistics 2*(1), 1-51.

Schertz, J. 2025. Individual uniformity in phonetic imitation: Assessing the stability of individual variability across features and tasks. *Journal of Phonetics 108*, 101376.

Schöner, G., Kopecz, K., & Erlhagen, W. 1997. The dynamic neural field theory of motor programming: Arm and eye movements. In *Advances in Psychology 119*, 271-310.

Schupack, D. A., Mars, R. A., Voelker, D. H., Abeykoon, J. P., & Kashyap, P. C. 2022. The promise of the gut microbiome as part of individualized treatment strategies. *Nature reviews Gastroenterology & hepatology 19*(1), 7-25.

Schweizer, D., Brunelle, M., Ahn, S., & Audet, A. 2025. A laryngeal and lingual ultrasound study of the Canadian French voicing contrast. *Journal of the International Phonetic Association 55*(1-2), 35-64.

Schweizer, D., Ahn, S., & Kwon, H. 2026. Acoustic properties of Seoul Korean stops in phrase-initial and word-medial positions. *Journal of the International Phonetic Association 56*, e6.

Schweitzer, A., Lewandowski, N., & Duran, D. 2017. Social attractiveness in dialogs. *Proceedings of Interspeech 2017*, 2243-2247.

Seyfarth, S., & Garellek, M. 2018. Plosive voicing acoustics and voice quality in Yerevan Armenian. *Journal of Phonetics 71*, 425-450.

Shantz, K. T. W. 2018. *Why wait? Psycholinguistic investigations of the roles of learning condition and gender stability in L2 gender-based anticipation*. Ph.D. dissertation, University of Illinois at Urbana-Champaign.

Shastri, U., Raj, K. K. P., Mathew, M., Kalaiah, M. K., & Uppunda, A. K. 2019. Relationship between working memory and identification of a few native phonetic contrasts. *Communication Sciences & Disorders 24*(1), 117-128.

Shih C, Möbius B, Narasimhan B. 1999. Contextual effects on consonantal voicing profiles: a cross-linguistic study. *Proceedings of the 14th International Congress of the Phonetic Sciences (ICPhS)*, 989-992.

Shockley, K., Sabadini, L., & Fowler, C. A. 2004. Imitation in shadowing words. *Perception & Psychophysics 66*(3), 422-429.

Shue, Y.-L., Keating, P., Vicenik, C. & Yu, K. 2011. VoiceSauce: A program for voice analysis. *Proceedings of the International 17th Congress of Phonetic Science,* 1846-1849.

Shultz, A. A., Francis, A. L., & Llanos, F. 2012. Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America 132*(2), EL95-EL101.

Siegler, R. S. 1994. Cognitive variability: A key to understanding cognitive development. *Current directions in psychological science 3*(1), 1-5.

Sloos, M. 2013. The reversal of the BÄREN-BEEREN merger in Austrian Standard German. *The Mental Lexicon 8*(3), 353-371.

Solé, M. J. 2018. Articulatory adjustments in initial voiced stops in Spanish, French and English. *Journal of Phonetics 66*, 217-241.

Stanley, J. A., & Sneller, B. 2023. Sample size matters in calculating Pillai scores. *The Journal of the Acoustical Society of America 153*(1), 54-67.

Stephani, T., Waterstraat, G., Haufe, S., Curio, G., Villringer, A., & Nikulin, V. V. 2020. Temporal signatures of criticality in human cortical excitability as probed by early somatosensory responses. *The Journal of Neuroscience 40*(34), 6572-6583.

Stevenson, S., & Zamuner, T. 2017. Gradient phonological relationships: Evidence from vowels in French. *Glossa: A journal of general linguistics 2*(1), 1-22.

Stone, M., Rizk, S., Woo, J., Murano, E. Z., Chen, H., & Prince, J. L. 2012. Frequency of apical and laminal/s/in normal and postglossectomy patients. *Journal of medical speech-language pathology 20*(4), 19.

Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., & Nishi, K. 2007. Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *The Journal of the Acoustical Society of America 122*(2), 1111-1129.

Streamer, J. E., & Queen, R. M. 2024. Movement variability and limb loading symmetry during simulated daily functional tasks. *Human Movement Science 98*, 103289.

Sumby, W. H., & Pollack, I. 1954. Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America 26*(2), 212-215.

Sundara, M. 2005. Acoustic-phonetics of coronal stops: A cross-language study of Canadian English and Canadian French. *The Journal of the Acoustical Society of America 118*(2), 1026-1037.

Sundara, M., Polka, L. & Baum, S. 2006. Production of coronal stops by simultaneous bilingual adults. *Bilingualism: Language and Cognition 9*(1), 97-114.

Tạ, T. T., Brunelle, M., & Nguyễn, T. Q. 2022. Voicing and register in Ngãi Giao Chrau: Production and perception studies. *Journal of Phonetics 90*, 101115.

Tabain, M., Kochetov, A., & Beare, R. 2020. An ultrasound and formant study of manner contrasts at four coronal places of articulation. *The Journal of the Acoustical Society of America 148*(5), 3195-3217.

Tan, X., Fan, S., Zheng, H., Liu, X., & Dai, C. 2026. Exploring Muscle Synergy Consistency and Physiological Interpretability in Speech Evaluation Using High-Density sEMG. *IEEE Transactions on Biomedical Engineering 73*(2), 923-933.

Tanner, J., Sonderegger, M., & Stuart-Smith, J. 2020. Structured speaker variability in Japanese stops: Relationships within versus across cues to stop voicing. *The Journal of the Acoustical Society of America 148*(2), 793-804.

Tennant, J. 2012. Laurentian French phonology in a majority setting outside Québec. In Gess, R., Lyche, C. & Meisenburg, T. (eds.) *Phonological Variation in French: Illustrations from three continents.* Amsterdam: John Benjamins, 313-339.

The Human Microbiome Project Consortium. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207-214.

Toda, M. 2009. *Étude articulatoire et acoustique des fricatives sibilantes.* Ph.D. dissertation, Université de la Sorbonne Nouvelle-Paris III.

Tremblay, P., & Small, S. L. 2011. On the context-dependent nature of the contribution of the ventral premotor cortex to speech perception. *NeuroImage 57*(4), 1561-1571.

Trevizan, R., Ferrero, V., Telles, F., Oliveira, L., Melo, L., Amorim, F.W. & Oliveira, P.E. 2026. Floral intra-individual variation differs between morphs and is linked to pollen deposition in self-incompatible distylous plants. *Plant Biology 28*(7), 2327-2336.

Trudeau-Fisette, P., Vidou, C., & Ménard, L. 2024. Speech sensorimotor relationships in francophone preschoolers and adults: Adaptation to real-time auditory feedback perturbations. *PLOS One 19*(8), e0306246.

Tuller, B., Nguyen, N., Lancia, L., & Vallabha, G. K. 2011. Nonlinear dynamics in speech perception. In Huys, R., Jirsa, V.K. (eds.) *Nonlinear Dynamics in Human Behavior. Studies in Computational Intelligence 328*. Berlin: Springer, 135-150.

Ulrich, N., Allassonnière-Tang, M., Pellegrino, F., & Dediu, D. 2021. Identifying the Russian voiceless non-palatalized fricatives /f/, /s/, and /ʃ/ from acoustic cues using machine learning. *The Journal of the Acoustical Society of America 150*(3), 1806-1820.

Ulrich, N., Pellegrino, F., & Allassonnière-Tang, M. 2023. Intra-and inter-speaker variation in eight Russian fricatives. *The Journal of the Acoustical Society of America 153*(4), 2285-2285.

Uluşahin, O., Bosker, H. R., McQueen, J. M., & Meyer, A. S. 2024. Knowledge of a talker's f0 affects subsequent perception of voiceless fricatives. *Proceedings of Speech Prosody 2024*, 432-436.

Urbán, T., Caballero, C., Barbado, D., & Moreno, F. J. 2019. Do intentionality constraints shape the relationship between motor variability and performance?. *PLOS One 14*(4), e0214237.

Vainio, L., Schulman, M., Tiippana, K., & Vainio, M. 2013. Effect of syllable articulation on precision and power grip performance. *PLOS One 8*(1), e53061.

Vouloumanos, A., & Werker, J. F. 2004. Tuned to the signal: the privileged status of speech for young infants. *Developmental science 7*(3), 270-276.

Walker, A., Hay, J., Drager, K., & Sanchez, K. 2018. Divergence in speech perception. *Linguistics 56*(1), 257-278.

Walker, D. C. 1984. *The pronunciation of Canadian French*. Ottawa: University of Ottawa Press.

Wang, N., & Feng, G. 2025. Driving factors of auditory category learning success. *Neuroscience & Biobehavioral Reviews*, 106435.

Wedel, A., Jackson, S., & Kaplan, A. 2013. Functional load and the lexicon: Evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change. *Language and speech*, *56*(3), 395-417.

Weirich, M., & Fuchs, S. 2013. Palatal morphology can influence speaker-specific realizations of phonemic contrasts. *Journal of Speech, Language, and Hearing Research 56*(6), 1894-1908.

Weirich, M., & Simpson, A. P. 2018. Individual differences in acoustic and articulatory undershoot in a German diphthong–Variation between male and female speakers. *Journal of Phonetics 71*, 35-50.

Werker, J. F., & Tees, R. C. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development 7*(1), 49-63.

Westbury, John R. 1983. Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America 73*(4), 1322-1336.

Westbury, J. R. 1994. *X-ray microbeam speech production database user's handbook*. Madison: University of Wisconsin.

Whalen, D. H. 1999. Three lines of evidence for direct links between production and perception in speech. In *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS),* 1257-1260.

Whalen, D. H., Chen, W. R., Tiede, M. K., & Nam, H. 2018. Variability of articulator positions and formants across nine English vowels. *Journal of Phonetics 68*, 1-14.

Winn, M. B., Rhone, A. E., Chatterjee, M., & Idsardi, W. J. 2013. The use of auditory and visual context in speech perception by listeners with normal hearing and listeners with cochlear implants. *Frontiers in psychology 4*, 824.

Winn, M. B. 2014a. Make_formant_continuum [Praat script]. Retrieved from: https://github.com/mwinn83/Praat/blob/master/Make_formant_continuum.txt

Winn, M. B. 2014b. Make_fricative_continuum [Praat script]. Retrieved from: https://github.com/mwinn83/Praat/blob/master/Make_fricative_continuum.txt

Winn, M. B., & Litovsky, R. Y. 2015. Using speech sounds to test functional spectral resolution in listeners with cochlear implants. *The Journal of the Acoustical Society of America 137*(3), 1430-1442.

Wu, H. G., Miyamoto, Y. R., Castro, L. N. G., Ölveczky, B. P., & Smith, M. A. 2014. Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nature neuroscience 17*(2), 312-321.

Yang, C., Bouffard, J., Srinivasan, D., Ghayourmanesh, S., Cantú, H., Begon, M., & Côté, J. N. 2018. Changes in movement variability and task performance during a fatiguing repetitive pointing task. *Journal of biomechanics 76*, 212-219.

Yu, A. 2019. On the nature of the perception-production link: Individual variability in English sibilant-vowel coarticulation. *Laboratory Phonology 10*(1), 1-29.

Yu, A. 2022. Perceptual cue weighting is influenced by the listener's gender and subjective evaluations of the speaker: the case of English stop voicing. *Frontiers in Psychology 13*, 840291.

Zaharchuk, H. A., Olson, H. E., Mechtenberg, H., Phillips, M. C., & Myers, E. B. 2026. Measuring age-related differences in phonetic gradiency with the visual analogue scale and eye tracking. *The Journal of the Acoustical Society of America 159*(3), 2460-2476.

Zellou, G., Scarborough, R., & Nielsen, K. 2016. Phonetic imitation of coarticulatory vowel nasalization. *The Journal of the Acoustical Society of America 140*(5), 3560-3575.

Zhai, W., Clayards, M., & Sonderegger, M. 2026. Are individual differences in lexical and coarticulatory biases linked across speech tasks?. *The Journal of the Acoustical Society of America 159*(1), 647-663.

Zhang, Y. & Goldstein, L. 2023. Stop voicing and devoicing as articulatory tasks: A cross-linguistic RT-MRI study. *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS)*, 1011-1015.

Zhang, Z. 2021. Contribution of laryngeal size to differences between male and female voice production. *The Journal of the Acoustical Society of America 150*(6), 4511-4521.

Zoefel, B., & Kösem, A. 2024. Neural tracking of continuous acoustics: properties, speech-specificity and open questions. *European Journal of Neuroscience 59*(3), 394-414.

Appendix: Coefficient tables

Table A1. Results of linear regression on closure duration for voiced and voiceless stops.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	99.527	3.002	54.9	33.149	< .001
Placecoronal	-7.451	1.095	14791.603	-6.807	< .001
Placevelar	-12.992	1.101	14791.789	-11.806	< .001
WordPosWord-medial	-16.924	2.555	65.597	-6.623	< .001
Voicingvoiceless	7.713	1.048	14791.599	7.358	< .001
Placecoronal:WordPosWord-medial	2.058	1.336	14791.422	1.54	.124
Placevelar:WordPosWord-medial	3.027	1.457	14791.313	2.078	.038
Placecoronal:Voicingvoiceless	-2.16	1.528	14791.502	-1.414	.158
Placevelar:Voicingvoiceless	-10.654	1.433	14798.054	-7.437	< .001
WordPosWord-medial:Voicingvoiceless	-1.227	1.407	14791.205	-0.872	.383
Placecoronal:WordPosWord-medial:Voicingvoiceless	41.833	1.875	14791.283	22.31	< .001
Placevelar:WordPosWord-medial:Voicingvoiceless	11.961	1.964	14794.651	6.091	< .001

Table A2. Results of linear regression on voicing proportion for voiced stops.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	86.485	1.861	61.265	46.462	< .001
Placecoronal	3.992	1.316	16.856	3.033	.008
Placevelar	-0.867	1.319	16.984	-0.657	.52
WordPosWord-medial	8.638	1.835	45.928	4.707	< .001
Placecoronal:WordPosWord-medial	-3.384	1.777	16.227	-1.905	.075
Placevelar:WordPosWord-medial	-3.26	1.783	16.456	-1.828	.086

Table A3. Results of linear regression on voicing proportion for voiceless stops.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	15.22	1.591	65.281	9.565	< .001
Placecoronal	-5.615	1.26	19.752	-4.456	< .001
Placevelar	-5.943	1.09	19.685	-5.45	< .001
WordPosWord-medial	-8.244	1.49	44.733	-5.534	< .001
Placecoronal:WordPosWord-medial	2.678	1.744	19.302	1.536	.141
Placevelar:WordPosWord-medial	0.605	1.626	19.227	0.372	.714

Table A4. Results of linear regression on voicing proportion for voiced stops in word-medial position.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	93.677	1.363	86.084	68.718	< .001
Placecoronal	1.879	1.032	2714.369	1.821	.069
Placevelar	-1.979	1.015	2713.912	-1.95	.051
Vowel/ɛ/	1.847	1.018	2713.937	1.814	.07
Vowel/u/	2.253	1.016	2713.923	2.217	.077
Vowel/o/	1.697	1.016	2713.922	1.67	.095
Placecoronal:Vowel/ɛ/	-1.65	1.45	2714.159	-1.138	.255
Placevelar:Vowel/ɛ/	-2.166	1.51	2715.07	-1.435	.151
Placecoronal:Vowel/u/	-2.689	1.449	2714.015	-1.856	.064
Placevelar:Vowel/u/	-2.359	1.439	2713.939	-1.639	.101
Placecoronal:Vowel/o/	-0.784	1.444	2714.125	-0.543	.587
Placevelar:Vowel/o/	-2.056	1.445	2714.101	-2.807	.075

Table A5. Results of linear regression on voicing proportion for voiceless stops in word-medial position.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	6.251	1.135	66.862	5.508	< .001
Placecoronal	-2.318	0.654	2752.093	-3.546	< .001
Placevelar	-5.171	0.655	2752.105	-7.899	< .001
Vowel/ɛ/	-1.389	0.659	2752.207	-3.017	.063
Vowel/o/	6.066	0.657	2752.085	9.238	< .001
Vowel/u/	-1.181	0.663	2752.203	-1.782	.075
Placecoronal:Vowel/ɛ/	1.461	0.928	2752.117	1.574	.116
Placevelar:Vowel/ɛ/	2.837	0.928	2752.139	3.058	.002
Placecoronal:Vowel/o/	-4.728	0.925	2752.064	-5.111	< .001

Placevelar:Vowel/o/	-5.597	0.931	2752.164	-6.012	< .001
Placecoronal:Vowel/u/	0.796	0.933	2752.087	0.853	.394
Placevelar:Vowel/u/	2.106	0.938	2752.123	2.246	.025

Table A6. Results of linear regression on post-release aspiration duration voiceless stops.

	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	0.322	0.03	221.719	10.614	< .001
Placecoronal	0.063	0.035	476.029	1.788	.074
Placevelar	0.121	0.035	476.029	3.399	.001
Voicingvoiceless	-0.061	0.041	275.593	-1.487	.138
WordPosWord-medial	-0.108	0.035	477.111	-3.051	.002
Placecoronal:Voicingvoiceless	-0.083	0.05	476.552	-1.655	.099
Placevelar:Voicingvoiceless	-0.103	0.05	476.552	-2.066	.039
Placecoronal:WordPosWord-medial	0.07	0.05	476.029	1.397	.163
Placevelar:WordPosWord-medial	0.104	0.05	476.029	2.081	.038
Voicingvoiceless:WordPosWord-medial	0.096	0.05	477.058	1.93	.054
Placecoronal:Voicingvoiceless:WordPosWord-medial	-0.123	0.07	476.293	-1.751	.081
Placevelar:Voicingvoiceless:WordPosWord-medial	-0.324	0.07	476.293	-4.605	< .001

Table A7. Results of linear regression on post-release aspiration duration for partially voiced *voiced* stops.

	Estimate	**Std. Error**	**df**	**t value**	**Pr(>|t|)**
(Intercept)	0.322	0.03	221.719	10.614	< .001
Placecoronal	0.063	0.035	476.029	1.788	.074
Placevelar	0.121	0.035	476.029	3.399	.001
Voicingvoiceless	-0.061	0.041	275.593	-1.487	.138
WordPosWord-medial	-0.108	0.035	477.111	-3.051	.002
Placecoronal:Voicingvoiceless	-0.083	0.05	476.552	-1.655	.099
Placevelar:Voicingvoiceless	-0.103	0.05	476.552	-2.066	.039

Placecoronal:WordPosWord-medial	0.07	0.05	476.029	1.397	.163
Placevelar:WordPosWord-medial	0.104	0.05	476.029	2.081	.038
Voicingvoiceless:WordPosWord-medial	0.096	0.05	477.058	1.93	.054
Placecoronal:Voicingvoiceless:WordPosWord-medial	-0.123	0.07	476.293	-1.751	.081
Placevelar:Voicingvoiceless:WordPosWord-medial	-0.324	0.07	476.293	-4.605	< .001

Table A8. Results of linear regression on post-release aspiration duration for partially voiced *voiced* stops and all voiceless stops.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	8.871	2.817	2153.551	3.149	.002
Placecoronal	0.657	2.645	6815.973	0.248	.804
Placevelar	12.245	2.508	6816.470	4.882	< .001
Voicingvoiceless	24.307	2.192	6817.693	11.090	< .001
Vowel/a/	-0.024	2.707	6816.336	-0.009	.993
Vowel/e/	-0.593	3.066	6816.189	-0.193	.847
Vowel/ɛ/	-0.250	3.153	6816.498	-0.079	.937
Vowel/o/	4.869	3.206	6816.099	1.519	.129
Vowel/u/	9.142	1.626	6817.249	5.621	< .001
Placecoronal:Voicingvoiceless	5.509	2.828	6815.828	1.948	.051
Placevelar:Voicingvoiceless	16.465	2.702	6816.221	6.093	< .001
Placecoronal:Vowel/a/	4.124	2.788	6815.948	1.479	.139
Placevelar:Vowel/a/	-0.998	2.716	6816.363	-0.367	.713
Placecoronal:Vowel/e/	5.407	3.615	6815.659	1.495	.135
Placevelar:Vowel/e/	2.931	3.265	6815.944	0.898	.369
Placecoronal:Vowel/ɛ/	1.003	3.580	6815.765	0.280	.779
Placevelar:Vowel/ɛ/	2.983	3.429	6816.201	0.870	.384
Placecoronal:Vowel/o/	-1.896	3.682	6815.677	-0.515	.607
Placevelar:Vowel/o/	-3.681	3.324	6816.063	-1.107	.268
Voicingvoiceless:Vowel/a/	-10.816	2.310	6816.851	-4.682	< .001
Voicingvoiceless:Vowel/e/	-12.094	2.783	6815.682	-4.346	< .001
Voicingvoiceless:Vowel/ɛ/	-13.282	2.880	6816.087	-4.612	< .001
Voicingvoiceless:Vowel/o/	-7.978	2.936	6815.519	-2.717	.007
Placecoronal:Voicingvoiceless:Vowel/a/	-2.698	3.014	6815.809	-0.895	.371
Placevelar:Voicingvoiceless:Vowel/a/	-5.244	2.936	6816.195	-1.786	.074
Placecoronal:Voicingvoiceless:Vowel/e/	0.288	3.877	6815.571	0.074	.941

Placevelar:Voicingvoiceless:Vowel/e/	2.389	3.555	6815.753	0.672	.502
Placecoronal:Voicingvoiceless:Vowel/ɛ/	1.583	3.845	6815.619	0.412	.681
Placevelar:Voicingvoiceless:Vowel/ɛ/	-6.943	3.707	6816.011	-1.873	.061
Placecoronal:Voicingvoiceless:Vowel/o/	-2.446	3.940	6815.591	-0.621	.535
Placevelar:Voicingvoiceless:Vowel/o/	-10.099	3.612	6815.878	-2.796	.005

Table A9. Results of linear regression on Coefficient of Variation of post-release aspiration duration in voiceless stops.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	0.477	0.014	105.766	33.500	< .001
Placecoronal	-0.143	0.015	96.000	-9.353	< .001
Placevelar	-0.184	0.015	96.000	-12.085	< .001

Table A10. Results of linear regression on Coefficient of Variation of voicing proportion in voiced stops.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	0.133	0.015	77.905	9.053	< .001
Placecoronal	-0.032	0.012	96.000	-2.652	.009
Placevelar	0.028	0.012	96.000	2.250	.027

Table A11. Results of linear regression on Coefficient of Variation for F1 in vowels.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	0.059	0.004	165.901	15.893	< .001
Vowel/ɛ/	0.008	0.005	144.000	1.841	.068
Vowel/o/	-0.002	0.005	144.000	-0.498	.619
Vowel/u/	0.015	0.005	144.000	3.247	.001

Table A12. Results of linear regression on Coefficient of Variation for F2 in vowels.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	0.046	0.003	173.90	10.071	< .001
Vowel/ɛ/	0.002	0.001	144.00	0.523	.602
Vowel/o/	0.09	0.001	144.00	17.634	< .001
Vowel/u/	0.156	0.001	144.00	27.909	< .001

Table A13. Results of linear regression on type of Coefficient of Variation (F1 and F2) in vowels.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	0.059	0.004	328.121	14.859	< .001
Vowel/ɛ/	0.008	0.005	336.000	1.645	.101
Vowel/o/	-0.002	0.005	336.000	-0.445	.656
Vowel/u/	0.015	0.005	336.000	2.902	.004
CoV_TypeCoVF2	-0.014	0.005	336.000	-2.818	.005
Vowel/ɛ/:CoV_TypeCoVF2	-0.006	0.007	336.000	-0.760	.448
Vowel/o/:CoV_TypeCoVF2	0.101	0.007	336.000	13.926	< .001
Vowel/u/:CoV_TypeCoVF2	0.138	0.007	336.000	19.056	< .001

Table A14. Results of linear regression on normalized F1 of vowels across the two recording sessions (Production A and Production B).

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	378.875	2.778	17.582	136.377	< .001
SessionProdB	2.453	4.484	14.875	0.547	.592
Vowel/ɛ/	159.795	2.523	3419.007	63.340	< .001
Vowel/o/	23.081	2.580	3420.154	8.945	< .001
Vowel/u/	-70.017	2.642	3423.984	-26.499	< .001
SessionProdB:Vowel/ɛ/	4.789	3.562	3418.665	1.344	.179
SessionProdB:Vowel/o/	2.551	3.646	3419.052	0.700	.484
SessionProdB:Vowel/u/	9.594	3.696	3422.415	2.595	.009

Table A15. Results of linear regression on normalized F2 of vowels across the two recording sessions (Production A and Production B).

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	2251.732	13.148	15.785	171.267	< .001
SessionProdB	-19.202	13.665	30.653	-1.405	.170
Vowel/ɛ/	-294.673	11.221	3425.687	-26.261	< .001
Vowel/o/	-1379.905	11.458	3427.183	-120.429	< .001
Vowel/u/	-1284.829	11.831	3430.827	-108.596	< .001
SessionProdB:Vowel/ɛ/	2.648	15.736	3426.005	0.168	.866
SessionProdB:Vowel/o/	64.047	16.123	3427.761	3.972	< .001
SessionProdB:Vowel/u/	31.028	16.475	3430.939	1.883	.060

Table A16. Results of linear regression on Euclidean Distance between vowels across the two recording sessions (Production A and Production B).

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	47.728	13.100	26.298	3.643	0.001
Vowel/ɛ/	-4.669	14.928	27.000	-0.313	.757
Vowel/o/	22.396	14.928	27.000	1.500	.145
Vowel/u/	10.306	14.928	27.000	0.690	.496

Table A17. Results of linear regression on Coefficient of Variation of F1 of vowels across the two recording sessions (Production A and Production B).

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	0.065	0.008	26.697	7.977	< .001
SessionProdB	-0.010	0.008	63.000	-1.168	.247
Vowel/ɛ/	-0.001	0.008	63.000	-0.139	.890
Vowel/o/	-0.008	0.008	63.000	-0.947	.347
Vowel/u/	0.009	0.008	63.000	1.068	.289
SessionProdB:Vowel/ɛ/	-0.003	0.012	63.000	-0.271	.787
SessionProdB:Vowel/o/	0.009	0.012	63.000	0.790	.432
SessionProdB:Vowel/u/	0.013	0.012	63.000	1.094	.278

Table A18. Results of linear regression on Coefficient of Variation of F2 of vowels across the two recording sessions (Production A and Production B).

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	0.047	0.011	69.505	4.290	< .001
SessionProdB	0.002	0.015	63.000	0.130	.897
Vowel/ɛ/	0.004	0.015	63.000	0.237	.813
Vowel/o/	0.105	0.015	63.000	7.062	< .001
Vowel/u/	0.168	0.015	63.000	11.321	< .001
SessionProdB:Vowel/ɛ/	-0.004	0.021	63.000	-0.209	.835
SessionProdB:Vowel/o/	-0.002	0.021	63.000	-0.107	.915
SessionProdB:Vowel/u/	0.012	0.021	63.000	0.565	.574

Table A19. Results of linear regression on normalized Center of Gravity of fricatives across the two recording sessions (Production A and Production B).

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	3862.571	47.049	742.000	82.097	< .001
SessionProdB	205.263	65.151	742.000	3.151	.002
Placedental	2596.082	66.081	742.000	39.286	< .001
SessionProdB:Placedental	-19.225	91.982	742.000	-0.209	.835

Table A20. Results of linear regression on Coefficient of Variation of Center of Gravity of fricatives across the two recording sessions (Production A and Production B).

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	0.056	0.009	22.166	6.044	< .001
ProdP2	0.014	0.010	27.000	1.448	.159
Placedental	0.038	0.010	27.000	3.930	.001
ProdP2:Placedental	-0.005	0.014	27.000	-0.391	.699

Table A21. Results of logistical regression on likelihood of VOICED responses for the stop voicing continuum in the perception experiment.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.575	0.157	16.351	< .001
Aspiration	-2.978	0.126	-23.644	< .001
Voicing	1.78	0.117	15.241	< .001
Placebilabial	-3.122	0.104	-30.014	< .001
Aspiration:Voicing	0.154	0.04	3.857	< .001
Aspiration:Placebilabial	0.592	0.085	6.985	< .001
Voicing:Placebilabial	0.353	0.065	5.432	< .001

Table A22. Results of linear regression on Coefficient of Unalikeability across the five continua in the perception experiment.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	0.11	0.011	227.785	9.66	< .001
TypeCoU_Fricative	-0.015	0.015	192	-0.961	.338
TypeCoU_u/o	0.01	0.015	192	0.68	.498
TypeLabialStop	0.012	0.015	192	0.784	.434
TypeVelarStop	0.008	0.015	192	0.498	.619