

Spoken word recognition in native and second language Canadian French: Phonetic detail
and representation of vowel nasalization

Félix Desmeules-Trudel

A thesis submitted in partial fulfillment of the requirements for the
Doctorate in Philosophy degree in Linguistics

Department of Linguistics
Faculty of Arts
University of Ottawa

Table of Contents

TABLE OF CONTENTS	II
LIST OF FIGURES	V
LIST OF TABLES	X
ABSTRACT.....	XII
RÉSUMÉ	XIII
ACKNOWLEDGEMENTS / REMERCIEMENTS	XIV
1 INTRODUCTION	1
1.1 Phonetic integration and recognition	5
1.2 Vowel nasalization in (Canadian) French.....	7
1.2.1 Articulation	8
1.2.2 Acoustic properties	11
1.3 Eye tracking and the visual world paradigm	13
2 NATIVE LANGUAGE PERCEPTION AND WORD RECOGNITION.....	17
2.1 Categorical perception of phonological contrasts	17
2.2 Gradience in speech sound processing	19
2.2.1 Other accounts of coarticulatory and gradient processing	21
2.3 Phonetics and word recognition.....	23
3 L1 PERCEPTION OF CANADIAN FRENCH VOWEL NASALITY AND NASAL COARTICULATION DURING SPOKEN WORD RECOGNITION	26
3.1 Experiment 1a	29
3.1.1 Methods.....	30
3.1.1.1 Participants.....	30
3.1.1.2 Stimuli and experimental conditions.....	32
3.1.1.3 Procedure	39
3.1.1.4 Analysis and variables	41
3.1.2 Results.....	46
3.1.2.1 Chosen images and reaction times	47
3.1.2.2 Eye movements	52
3.1.3 Additional results and discussion.....	64
3.2 Experiment 1b	67
3.2.1 Methods.....	68

3.2.1.1	Participants	68
3.2.1.2	Stimuli	69
3.2.1.3	Procedure and analysis	70
3.2.2	Results	71
3.2.2.1	Chosen images and reaction times	71
3.2.2.2	Eye movements	78
3.2.3	Additional results and discussion	89
3.3	General discussion.....	93
4	<u>SECOND LANGUAGE SPEECH PERCEPTION AND WORD RECOGNITION.....</u>	<u>96</u>
4.1	Naïve speech perception	98
4.1.1	The Perceptual Assimilation Model.....	98
4.1.2	Experimental evidence in favour of PAM classification	100
4.2	Second language speech perception.....	103
4.2.1	The Speech Learning Model	103
4.2.2	The Perceptual Assimilation Model for second language perception.....	104
4.3	Lexical recognition in a second language.....	107
4.4	Language exposure and second language acquisition	110
4.5	English vowel nasalization and comparison with Canadian French.....	113
5	<u>L2 PERCEPTION OF CANADIAN FRENCH VOWEL NASALIZATION BY</u>	
	<u>NATIVE ENGLISH LISTENERS DURING SPOKEN WORD RECOGNITION</u>	<u>115</u>
5.1	Experiment 2a	117
5.1.1	Methods.....	117
5.1.1.1	Participants	117
5.1.1.2	Stimuli and experimental conditions.....	120
5.1.1.3	Procedure	123
5.1.1.4	Analysis and variables	123
5.1.2	Results	126
5.1.2.1	Chosen images and reaction times	126
5.1.2.2	Eye movements	134
5.1.3	Additional results and discussion.....	147
5.2	Experiment 2b	151
5.2.1	Methods.....	151
5.2.1.1	Participants	151
5.2.1.2	Stimuli	153
5.2.1.3	Procedure and analysis.....	154
5.2.2	Results	154

5.2.2.1	Chosen image and reaction times.....	154
5.2.2.2	Eye movements	162
5.2.3	Additional results and discussion.....	174
5.3	General discussion.....	177
6	<u>GENERAL DISCUSSION AND CONCLUSION</u>	<u>181</u>
6.1	Summary of the results.....	181
6.2	Integrating L1 and L2 spoken word recognition	184
6.3	Phono-lexical representations include phonetic details	187
6.4	Limitations and further research	189
	<u>REFERENCES.....</u>	<u>193</u>
	<u>APPENDICES.....</u>	<u>213</u>
	Appendix 1	213
	Appendix 2	217
	Appendix 3	222
	Appendix 4	256

List of Figures

<i>Figure 1.1.</i> Nasal airflow patterns of nasal appendices compared to nasal consonants in closed and open syllables, from Desmeules-Trudel & Brunelle (2018)	11
<i>Figure 1.2.</i> Spectrogram of the vowel [ɛ̃] (word <i>vin</i> [vɛ̃] ‘wine’) pronounced by a female speaker of CF, and acoustic correlates of vowel nasalization.....	12
<i>Figure 2.1.</i> Participants’ %fix patterns to the <i>beach</i> or to the <i>peach</i> on a 0-40 ms VOT continuum, from McMurray et al. (2008).....	20
<i>Figure 3.1.</i> Possible predicted recognition patterns – Categorical or gradient.....	29
<i>Figure 3.2.</i> Distribution of self-reported L2 English proficiency scores out of 10 for Exp. 1a participants	31
<i>Figure 3.3.</i> Spectrogram of <i>dame/dent</i> [dam]/[dã] ‘lady/tooth’ stimulus in the 50N %NASDUR – Exp. 1a	35
<i>Figure 3.4.</i> Example of one experimental display for the <i>pain</i> (C $\tilde{V}$) – <i>peigne</i> (CVN) pair along with <i>pêche</i> (CVC) and <i>pluie</i> (unrelated).....	38
<i>Figure 3.5.</i> Proportions of image choices for nasal (C $\tilde{V}$) and coarticulated (CVN) images and GAMM predictions for image choice analysis (i.e. probability of C $\tilde{V}$ image choice by %NASDUR) – Exp. 1a	48
<i>Figure 3.6.</i> RT averages for nasal (C $\tilde{V}$) image choices by %NASDUR and coarticulated (CVN) image choices by %NASDUR – Exp. 1a	49
<i>Figure 3.7.</i> Predicted data by %NASDUR and CHOSEN IMAGE – Exp. 1a RTs	52
<i>Figure 3.8.</i> Proportions of fixations to the nasal (C $\tilde{V}$) image and coarticulated (CVN) image, nasal (C $\tilde{V}$) choices and coarticulated (CVN) choices – Exp. 1a	54
<i>Figure 3.9.</i> Predicted data by %NASDUR – Exp. 1a elog %fix to C $\tilde{V}$, C $\tilde{V}$ choices	58
<i>Figure 3.10.</i> Predicted data by CHOSEN IMAGE through TIME – Exp. 1a elog %fix to C $\tilde{V}$, C $\tilde{V}$ choices.....	59
<i>Figure 3.11.</i> Predicted data by %NASDUR and VOWEL QUALITY through TIME – Exp. 1a elog %fix to C $\tilde{V}$, C $\tilde{V}$ choices	60
<i>Figure 3.12.</i> Predicted data by %NASDUR – Exp. 1a elog %fix to CVN, CVN choices	62
<i>Figure 3.13.</i> Predicted data by %NASDUR and VOWEL QUALITY through TIME – Exp. 1a elog %fix to CVN, CVN choices.....	63
<i>Figure 3.14.</i> Distribution of self-reported L2 English proficiency scores out of 10 for Exp. 1b participants	69

<i>Figure 3.15.</i> Spectrogram of <i>dame/dent</i> [dam]/[dã] ‘lady/tooth’ stimulus in the 50N %NASDUR with no nasal appendix – Exp. 1b	70
<i>Figure 3.16.</i> Proportions of image choices for nasal (C $\tilde{V}$) and coarticulated (CVN) images and GAMM predictions for image choice analysis (i.e. probability of C $\tilde{V}$ image choice by %NASDUR) – Exp. 1b	72
<i>Figure 3.17.</i> RT averages for nasal (C $\tilde{V}$) image choices by %NASDUR and coarticulated (CVN) image choices by %NASDUR – Exp. 1	74
<i>Figure 3.18.</i> Predicted data by %NASDUR – Exp. 1b RTs	77
<i>Figure 3.19.</i> Predicted data by %NASDUR and VOWEL QUALITY – Exp. 1b RTs	78
<i>Figure 3.20.</i> Proportions of fixations to the nasal (C $\tilde{V}$) image and coarticulated (CVN) image, nasal (C $\tilde{V}$) choices and coarticulated (CVN) choices – Exp. 1b	79
<i>Figure 3.21.</i> Predicted data by %NASDUR – Exp. 1b <i>e</i> log %fix to C $\tilde{V}$, C $\tilde{V}$ choices	83
<i>Figure 3.22.</i> Predicted data by %NASDUR through TIME – Exp. 1b <i>e</i> log %fix to C $\tilde{V}$, C $\tilde{V}$ choices	84
<i>Figure 3.23.</i> Predicted data by %NASDUR and VOWEL QUALITY through TIME – Exp. 1b <i>e</i> log %fix to C $\tilde{V}$, C $\tilde{V}$ choices	85
<i>Figure 3.24.</i> Predicted data by VOWEL QUALITY through TIME – Exp. 1b <i>e</i> log %fix to CVN, CVN choices.....	88
<i>Figure 3.25.</i> Predicted data by %NASDUR and VOWEL QUALITY through TIME – Exp. 1b <i>e</i> log %fix to CVN, CVN choices.....	89
<i>Figure 4.1.</i> Illustration of the acoustic proximity (F1/F2) of Canadian French and Canadian English [ɛ]-[æ] vowels compared to Peruvian Spanish [e]-[a] vowels, reproduced from Escudero and Vasiliev (2011, p. EL278)	102
<i>Figure 5.1.</i> Possible predicted recognition patterns – Categorical or gradient.....	116
<i>Figure 5.2.</i> Distribution of self-reported L2 French proficiency scores out of 10 for Exp. 2a participants	119
<i>Figure 5.3.</i> Average proportion of current exposure (receptive) to and use (productive) of L2 French in various situations for Exp. 2a participants.....	120
<i>Figure 5.4.</i> Proportions of image choices for nasal (C $\tilde{V}$) and coarticulated (CVN) images and GAMM predictions for image choice analysis (i.e. probability of C $\tilde{V}$ image choice by %NASDUR) – Exp. 2a	128
<i>Figure 5.5.</i> RT averages for selected nasal (C $\tilde{V}$) images by %NASDUR and coarticulated (CVN) images by %NASDUR – Exp. 2a	130
<i>Figure 5.6.</i> Predicted data by EARLY-EXPOSURE-GROUP – Exp. 2a RTs.....	133
<i>Figure 5.7.</i> Predicted data by %NASDUR – Exp. 2a RTs	134

<i>Figure 5.8. Proportions of fixations to the nasal (C$\tilde{V}$) image and coarticulated (CVN) image, nasal (C$\tilde{V}$) choices and coarticulated (CVN) choices – Exp. 2a</i>	135
<i>Figure 5.9. Predicted data by EXPOSURE group – Exp. 2a elog %fix to C$\tilde{V}$, C$\tilde{V}$ choices</i>	139
<i>Figure 5.10. Predicted data %NASDUR – Exp. 2a elog %fix to C$\tilde{V}$, C$\tilde{V}$ choices</i>	140
<i>Figure 5.11. Predicted data by %NASDUR and VOWEL QUALITY through TIME – Exp. 2a elog %fix to C$\tilde{V}$, C$\tilde{V}$ choices</i>	141
<i>Figure 5.12. Predicted data by EXPOSURE – Exp. 2a elog %fix to CVN, CVN choices</i>	143
<i>Figure 5.13. Predicted data by %NASDUR – Exp. 2a elog %fix to CVN, CVN choices</i>	144
<i>Figure 5.14. Predicted data by VOWEL QUALITY through TIME – Exp. 2a elog %fix to CVN, CVN choices</i>	145
<i>Figure 5.15. Predicted data by VOWEL QUALITY and %NASDUR through TIME – Exp. 2a elog %fix to CVN, CVN choices</i>	146
<i>Figure 5.16. Distribution of self-reported L2 French proficiency scores out of 10 for Exp. 2b participants</i>	152
<i>Figure 5.17. Average proportion of current exposure (receptive) to and use (productive) of L2 French in various situations for Exp. 2b participants</i>	153
<i>Figure 5.18. Proportions of image choices for nasal (C$\tilde{V}$) and coarticulated (CVN) images and GAMM predictions for image choice analysis (i.e. probability of C$\tilde{V}$ image choice by %NASDUR) – Exp. 2b</i>	156
<i>Figure 5.19. RT averages for nasal (C$\tilde{V}$) image choices by %NASDUR and coarticulated (CVN) image choices by %NASDUR – Exp. 2b</i>	158
<i>Figure 5.20. Predicted data by %NASDUR – Exp. 2b RTs</i>	161
<i>Figure 5.21. Predicted data by %NASDUR and VOWEL QUALITY – Exp. 2b RTs</i>	162
<i>Figure 5.22. Proportions of fixations to the nasal (C$\tilde{V}$) image and coarticulated (CVN) image, nasal (C$\tilde{V}$) choices and coarticulated (CVN) choices – Exp. 2b</i>	164
<i>Figure 5.23. Predicted data by %NASDUR – Exp. 2b elog %fix to C$\tilde{V}$, C$\tilde{V}$ choices</i>	168
<i>Figure 5.24. Predicted data by VOWEL QUALITY through TIME – Exp. 2b elog %fix to C$\tilde{V}$, C$\tilde{V}$ choices</i>	169
<i>Figure 5.25. Predicted data by CHOSEN IMAGE through TIME – Exp. 2b elog %fix to C$\tilde{V}$, C$\tilde{V}$ choices</i>	170
<i>Figure 5.26. Predicted data by %NASDUR and VOWEL QUALITY through TIME – Exp. 2b elog %fix to C$\tilde{V}$, C$\tilde{V}$ choices</i>	171
<i>Figure 5.27. Predicted data by %NASDUR – Exp. 2b elog %fix to CVN, CVN choices</i>	173
<i>Figure 5.28. Predicted data by %NASDUR and VOWEL QUALITY through TIME – Exp. 2b elog %fix to CVN, CVN choices</i>	174
<i>Figure A.1. Predicted data by TRIAL NUMBER – Exp. 1a RTs</i>	222

<i>Figure A.2.</i> Predicted data by L2 (ENGLISH) PROFICIENCY score – Exp. 1a RTs.....	223
<i>Figure A.3.</i> Predicted data through TIME – Exp. 1a elog %fix to C $\tilde{V}$, C $\tilde{V}$ choices	224
<i>Figure A.4.</i> Predicted data by FREQUENCY (NASAL) – Exp. 1a elog %fix to C $\tilde{V}$, C $\tilde{V}$ choices	225
<i>Figure A.5.</i> Predicted data by TRIAL NUMBER through TIME – Exp. 1a elog %fix to C $\tilde{V}$, C $\tilde{V}$ choices.....	226
<i>Figure A.6.</i> Predicted data through TIME – Exp. 1a elog %fix to CVN, CVN choices.....	227
<i>Figure A.7.</i> Predicted data by TRIAL NUMBER – Exp. 1a elog %fix to CVN, CVN choices	228
<i>Figure A.8.</i> Predicted data by PROVINCE OF ORIGIN through TIME – Exp. 1a elog %fix to CVN, CVN choices	229
<i>Figure A.9.</i> Predicted data by TRIAL NUMBER through TIME – Exp. 1a elog %fix to CVN, CVN choices.....	230
<i>Figure A.10.</i> Predicted data by FREQUENCY (NASAL) through TIME – Exp. 1a elog %fix to CVN, CVN choices	231
<i>Figure A.11.</i> Predicted data through TIME – Exp. 1b elog %fix to C $\tilde{V}$, C $\tilde{V}$ choices	232
<i>Figure A.12.</i> Predicted data by TRIAL NUMBER through TIME – Exp. 1b elog %fix to C $\tilde{V}$, C $\tilde{V}$ choices.....	233
<i>Figure A.13.</i> Predicted data by TRIAL NUMBER – Exp. 1b elog %fix to CVN, CVN choices.....	234
<i>Figure A.14.</i> Predicted data by FREQUENCY (COARTICULATED) – Exp. 1b elog %fix to CVN, CVN choices.....	235
<i>Figure A.15.</i> Predicted data by TRIAL NUMBER – Exp. 1b elog %fix to CVN, CVN choices.....	236
<i>Figure A.16.</i> Predicted data by TRIAL NUMBER – Exp. 1b elog %fix to CVN, CVN choices.....	237
<i>Figure A.17.</i> Predicted data by (LOG) FREQUENCY (COARTICULATED) – Exp. 2a RTs.....	238
<i>Figure A.18.</i> Predicted data by TRIAL NUMBER – Exp. 2a RTs	239
<i>Figure A.19.</i> Predicted data through TIME – Exp. 2a elog %fix to C $\tilde{V}$, C $\tilde{V}$ choices	240
<i>Figure A.20.</i> Predicted data by TRIAL NUMBER – Exp. 2a elog %fix to C $\tilde{V}$, C $\tilde{V}$ choices.....	241
<i>Figure A.21.</i> Predicted data by TRIAL NUMBER through TIME – Exp. 2a elog %fix to C $\tilde{V}$, C $\tilde{V}$ choices.....	242
<i>Figure A.22.</i> Predicted data through TIME – Exp. 2a elog %fix to CVN, CVN choices	243
<i>Figure A.23.</i> Predicted data by TRIAL NUMBER – Exp. 2a elog %fix to CVN, CVN choices.....	244
<i>Figure A.24.</i> Predicted data by FREQUENCY (COARTICULATED) – Exp. 2a elog %fix to CVN, CVN choices.....	245
<i>Figure A.25.</i> Predicted data by TRIAL NUMBER through TIME – Exp. 2a elog %fix to CVN, CVN choices.....	246
<i>Figure A.26.</i> Predicted data by FREQUENCY (NASAL) through TIME – Exp. 2a elog %fix to CVN, CVN choices	247
<i>Figure A.27.</i> Predicted data by TRIAL NUMBER – Exp. 2b RTs.....	248

<i>Figure A.28.</i> Predicted data by TRIAL NUMBER – Exp. 2b elog %fix to C $\tilde{V}$, C $\tilde{V}$ choices.....	249
<i>Figure A.29.</i> Predicted data by TRIAL NUMBER – Exp. 2b elog %fix to C $\tilde{V}$, C $\tilde{V}$ choices.....	250
<i>Figure A.30.</i> Predicted data by TRIAL NUMBER through TIME – Exp. 2b elog %fix to C $\tilde{V}$, C $\tilde{V}$ choices.....	251
<i>Figure A.31.</i> Predicted data through TIME – Exp. 2b elog %fix to CVN, CVN choices	252
<i>Figure A.32.</i> Predicted data by FREQUENCY (COARTICULATED) – Exp. 2b elog %fix to CVN, CVN choices.....	253
<i>Figure A.33.</i> Predicted data by TRIAL NUMBER through TIME – Exp. 2b elog %fix to CVN, CVN choices.....	254
<i>Figure A.34.</i> Predicted data by FREQUENCY (COARTICULATED) through TIME – Exp. 2b elog %fix to CVN, CVN choices.....	255

List of Tables

<i>Table 1.1.</i> Minimal pairs of oral vowels and nasal vowels	8
<i>Table 3.1.</i> Word list (experimental stimuli).....	33
<i>Table 3.2.</i> Summary of the experimental conditions (%NASDUR) – Exp. 1a.....	35
<i>Table 3.3.</i> Auditory stimuli details	36
<i>Table 3.4.</i> GAMM results – Exp. 1a image choices	49
<i>Table 3.5.</i> Summary of the independent variables for the full RT GAMM – Exp. 1a	50
<i>Table 3.6.</i> GAMM results – Exp. 1a RTs	51
<i>Table 3.7.</i> Summary of the independent variables for the full GAMMs – Exp. 1a %fix	55
<i>Table 3.8.</i> GAMM results – Exp. 1a %fix to C $\tilde{V}$, C $\tilde{V}$ choices.....	56
<i>Table 3.9.</i> GAMM results – Exp. 1a %fix to CVN, CVN choices	61
<i>Table 3.10.</i> Summary of the experimental conditions (%NASDUR) – Exp. 1b	70
<i>Table 3.11.</i> GAMM results – Exp. 1b image choices.....	73
<i>Table 3.12.</i> Summary of the independent variables for the full RT GAMM – Exp. 1b	75
<i>Table 3.13.</i> GAMM results – Exp. 1b RTs	76
<i>Table 3.14.</i> Summary of the independent variables for the full GAMM – Exp. 1b %fix	81
<i>Table 3.15.</i> GAMM results – Exp. 1b %fix to C $\tilde{V}$, C $\tilde{V}$ choices.....	82
<i>Table 3.16.</i> GAMM results – Exp. 1b %fix to CVN, CVN choices.....	86
<i>Table 4.1.</i> Summary of the PAM-L2 predictions for discrimination and perceptual learning abilities in an L2	106
<i>Table 5.1.</i> Summary of the experimental conditions (%NASDUR) – Exp. 2a.....	121
<i>Table 5.2.</i> Word list (experimental stimuli).....	122
<i>Table 5.3.</i> GAMM results – Exp. 1a image choices	129
<i>Table 5.4.</i> Summary of the independent variables for the full RT GAMM – Exp. 2a	131
<i>Table 5.5.</i> GAMM results – Exp. 2a RTs	132
<i>Table 5.6.</i> Summary of the independent variable for the full GAMM – Exp. 2a %fix	137
<i>Table 5.7.</i> GAMM results – Exp. 2a %fix to C $\tilde{V}$, C $\tilde{V}$ choices.....	138
<i>Table 5.8.</i> GAMM results – Exp. 2a %fix to CVN, CVN choices	142
<i>Table 5.9.</i> GAMM results – Exp. 2b image choices.....	157
<i>Table 5.10.</i> Summary of the independent variables for the full RT GAMM – Exp. 2b	159
<i>Table 5.11.</i> GAMM results – Exp. 2b RTs.....	160
<i>Table 5.12.</i> Summary of the independent variable for the full GAMM – Exp. 2b %fix	166

<i>Table 5.13.</i> GAMM results – Exp. 2b %fix to C $\tilde{V}$, C $\tilde{V}$ choices.....	167
<i>Table 5.14.</i> GAMM results – Exp. 2b %fix to CVN, CVN choices.....	172
<i>Table 6.1.</i> Summary of the experimental results on the patterns of phonetic integration	184
<i>Table A.1.</i> Results of the L2 familiarisation phase	256

Abstract

Research has shown that fine-grained consonantal phonetic information can be gradiently integrated during spoken word recognition in the L1. However, the way listeners categorize vocalic phonetic information has not been investigated as thoroughly. Furthermore, second language (L2) listeners' processing of fine-grained information is not as well known as L1 processing.

L1 Canadian French (CF) listeners and L2 listeners (native English) were tested in an eye tracking paradigm with words containing partially nasalized (CVN) and fully nasal (C $\tilde{V}$) vowels. Stimuli were designed to have variable nasalization duration on the vowel, and sometimes include a short nasal consonant word-finally. The main goals were to determine how nasalization duration influences word recognition in an L1 and an L2, and if variations in phonetic details are gradiently or categorically integrated.

Results show that L1 listeners *gradiently* were able to identify the stimuli when they contained mismatching phonetic cues, while L2 listeners display more categorical patterns of recognition. When stimuli do not have conflicting phonetic cues, L1 listeners mostly identify words as C $\tilde{V}$, except when the vowel is not nasalized. For L2 listeners, the pattern was similar, but the rate of stimuli identification as phonological nasal (C $\tilde{V}$) was lower due to L1 transfer.

These results support the hypothesis that L1 listeners have phonological representations that include fine-grained phonetic information and that they consider it when recognizing words. On the other hand, L2 listeners who have less experience in the L2 display more categorical recognition patterns, probably because their representations include coarser phonetic information or because they cannot access fine-grained representations, given the cognitive demands of L2 processing. When words do not contain conflicting phonetic cues, patterns of recognition of both L1 and L2 listeners seem more categorical, even though L2 listeners displayed lower rates of identification than L1 listeners overall. This uncertainty can also be due to less detailed phonological representations or to their inability to access all the necessary information to recognize words. Overall, these results suggest that fine-grained phonetic information gradiently impacts word recognition, that it is part of phono-lexical representations, and that L2 processing is qualitatively and quantitatively different from L1 processing.

Résumé

La recherche a démontré que l'information phonétique fine des consonnes peut être intégrée de façon *gradiente* pour la reconnaissance de mots auditifs dans la langue maternelle (L1). Toutefois, la façon dont les locuteurs catégorisent l'information vocalique n'a pas été examinée aussi minutieusement, selon ma connaissance. De plus, le traitement de l'information phonétique fine par les locuteurs d'une langue seconde (L2) n'a pas été aussi détaillé que celui du traitement L1.

Des locuteurs L1 et L2 (de langue maternelle anglaise) du français canadien (FC) ont complété une étude oculométrique s'attardant au traitement des mots contenant des voyelles nasalisées coarticulatoires (CVN) ou nasales phonologiques (C $\tilde{V}$). Les stimuli contenaient des variations de durée de nasalisation sur les voyelles, et parfois une courte consonne nasale en fin de mot. Les buts principaux étaient de déterminer de quelle façon la durée de nasalisation influence la reconnaissance lexicale en L1 et L2, et si les variations des détails phonétiques sont intégrées de façon catégorique ou *gradiente*.

Les résultats montrent que les locuteurs L1 ont pu identifier les stimuli de façon gradiente lorsqu'ils contenaient des indices phonétiques contradictoires, alors que les locuteurs L2 ont présenté des patrons de reconnaissances plus catégoriques. Lorsque les stimuli ne contenaient pas d'indices contradictoires, les locuteurs L1 ont majoritairement identifié les stimuli comme C $\tilde{V}$, sauf si la voyelle n'était pas nasalisée du tout. En L2, les locuteurs ont présenté un patron similaire, mais la proportion d'identification en tant que C $\tilde{V}$ était plus basse, probablement étant donné le transfert L1.

Ces résultats sont cohérents avec l'hypothèse selon laquelle les locuteurs L1 ont des représentations phonologiques qui incluent l'information phonétique fine et qu'ils sont capables de considérer ce type d'information pour la reconnaissance de mots auditifs. Cependant, les locuteurs L2 qui ont moins d'expérience dans la langue-cible présentent des patrons plus catégoriques, probablement étant donné que leurs représentations incluent seulement de l'information phonétique plus grossière, ou parce qu'ils ne sont pas en mesure d'accéder à l'information plus fine à cause des demandes cognitives du traitement L2. De plus, lorsque les mots ne contiennent pas d'information contradictoire, les patrons de reconnaissance dans la L1 et la L2 semblent plutôt catégoriques, même si les locuteurs L2 avaient des proportions d'identification généralement plus basses que les locuteurs L1. Cette incertitude est probablement causée par l'absence de détails dans les représentations phonologiques ou à l'incapacité d'accéder à toute l'information nécessaire pour la reconnaissance des mots. En somme, ces résultats suggèrent que l'information phonétique fine a un impact gradient sur la reconnaissance lexicale, qu'elle fait partie des représentations phono-lexicales, et que le traitement en L2 est qualitativement et quantitativement différent du traitement en L1.

Acknowledgements / Remerciements

This dissertation work was made possible thanks to the astonishing support and dedication of Dr. Tania Zamuner throughout my doctoral journey. I could not have dreamed of a better mentor and supervisor – thanks for the opportunities and shared knowledge, intellectual rigor, and great understanding of what it is to complete grad school along with the support throughout. I am also very grateful to Dr. Marc Brunelle, who was a great co-author/mentor, and whose patience and support in me learning how to collect nasal airflow data (and report it!) was *on-ne-peut-plus* appreciated. Dr. Laura Sabourin also let me spy on her lab meetings for a couple of years, thanks for giving me the opportunity to learn and laugh with you and your ERPLings. I would also like to thank Dr. Ian MacKay for his great help and comments on some work I have done for the program, and Dr. Kevin McMullin for very insightful comments and advice on the current work. Finally, I am very grateful to Dr. Bob McMurray for agreeing to read and comment on my dissertation, it is highly appreciated and I learned a lot from your insights on my work.

I also want to thank my colleague-grad-schoolers from the Department, those with whom I managed through easier and tougher times (even syntax class), especially Suzanne Robillard, J-C Leclerc, Gita Zareikar, Santa Vĩnerte, Myriam Ducos, and all the others whose path I crossed in the last five years. I wish all of you guys the best of luck in your endeavors, you deserve it. Thank you also to all of the other professors and staff at uOttawa Linguistics, you all contributed to making my journey here a nice ride. I am also lucky to have met and discussed with the bright minds at the Centre for Child Language Research, the uOttawa Living Lab, the Sound Patterns Laboratory and the ERPLing Lab. It was a true privilege to meet all of you. I am greatly indebted to Émilie Piché for her awesome help with data collection for my dissertation.

Ma famille et mes amis ont été d'un support incroyable tout au long de mon parcours. Merci infiniment à mes parents pour leurs encouragements constants et inconditionnels, et à Dillon pour le quotidien, l'écoute et l'acceptation des à-côtés d'un doctorat. Salut à celles et ceux que j'ai rencontrés et avec qui je me suis lié d'amitié dans la région. L'Ontario français et l'ouest du Québec sont chanceux de vous avoir! À Charles et Camille, pour votre intérêt et votre curiosité, et pour vos visites dans les dernières années, je vous remercie. Quel bonheur d'avoir une fratrie comme la nôtre.

Finally, I would like to acknowledge the contribution of the University of Ottawa, the Fonds Québécois de recherche – Société et culture, as well as the Social Sciences and Humanities Research Council of Canada for the financial support that was provided to me in the last five years. I would also like to thank the professors who hired me to work on specific projects as a research assistant, especially Dr. Tania Zamuner and Dr. Marc Brunelle. The funding from which I have benefited enabled me to fully focus on my research and to gain wonderful insights by attending great conferences and trainings in Europe and North America.

Pour

Paul, le RÊVEUR

Anne-Marie, la FOUGUEUSE

René, le TRAVAILLANT

Denise, la MINUTIEUSE

Vous m'inspirez

Understanding the relation between phonetics and phonology is thus ultimately a matter of understanding the dual nature of the sign, and much of the difficulty in defining and delimiting their respective realms is ultimately due to the difficulty of deciding *what sort of abstractions we are dealing with when we study language*.
(emphasis mine; Ladd, 2011, p. 349)

1 Introduction

The relationship between phonetics and phonology has been a topic of debate for decades within linguistics. Traditionally, each field analyzes different objects: phonetics is the study of speech *sounds*, and phonology is the study of the language's sound *systems*. Consequently, researchers have often considered that investigating a language's phonology can be done independently from the phonetics, and that phonetic analysis does not always necessitate a detailed account of a language's phonology. However, other researchers have questioned these assumptions and considered the idea of linking phonetics and phonology for a long time. Even in the seminal *Sound Pattern of English*, Chomsky and Halle (1968) dedicate an entire chapter to the fact that phonetics and phonology have to be linked to a certain extent, at least in production. In the last thirty years, fields such as laboratory phonology, psycholinguistics, and theoretical phonology have shown growing interest in the interface between phonetics and phonology, considering sound change and language evolution (Garrett & Johnson, 2013; Kingston, 2007; Ohala, 2005), speech perception (Hume & Johnson, 2001; Mielke, 2003), speech and lexical processing (Ganong, 1980; McQueen, Cutler, & Norris, 2006; Newman, Sawusch, & Luce, 1997; Pierrehumbert, 2002), gradient phonological relationships (Cohn, 2006; Hall, 2013; Stevenson & Zamuner, 2017) and theories of phonological storage that include fine-grained phonetic information (McMurray, Tanenhaus, & Aslin, 2002; Toscano, McMurray, Dennhardt, & Luck, 2010). These recent advances have enabled researchers to focus on the phonetic content of phonological representations. From a psycholinguistic point of view, this dissertation aims to

document the link between phonetics and phonology, considering the question of the phonetic content of phonological representation and phonetic processing during spoken word recognition.

One way of investigating the phonetics-phonology interface is by looking at phonetic detail processing. Perception and use of phonetic detail, which can be seen as the variable ways a phoneme can be realized (e.g. variability in VOT for voiceless stops in English, amount and duration of coarticulatory vowel nasalization, amplitude and velocity of formant transitions), for word recognition have been studied, and this fine-grained information was shown to influence the retrieval of abstract phonemic and lexical information (Beddor, McGowan, Boland, Coetzee, & Brasher, 2013; Dahan, Magnuson, Tanenhaus, & Hogan, 2001b; Gow, 2003; McMurray, Clayards, Tanenhaus, & Aslin, 2008; McMurray et al., 2002; McMurray, Tanenhaus, & Aslin, 2009; Paquette-Smith, Fecher, & Johnson, 2016; Zamuner, Moore, & Desmeules-Trudel, 2016). For example, it has been shown that English listeners pay attention to vowel transitions towards a word-final consonant (Dahan et al., 2001b), to vowel nasalization (Beddor et al., 2013; Paquette-Smith et al., 2016; Zamuner, et al., 2016) or to voice onset time (McMurray et al., 2002; 2008) during spoken word recognition. The current work presents additional results on the influence of variations in phonetic details and low-level acoustic information on lexical processing by looking at differences in vowel nasalization timing in Canadian French (CF), a language that contains both phonologically nasal ($\tilde{V}$) and coarticulatorily nasalized (VN) vowels.

The studies on the use of coarticulation and phonetic variations cited above (Dahan et al., 2001b; Paquette-Smith et al., 2016; Zamuner et al., 2016) also point towards a theory of phonological-lexical storage that includes fine-grained coarticulatory information into representations, as well as the gradient use of phonetic cues during word recognition (McMurray et al., 2002; McMurray et al., 2008). Previous results from the literature suggest that listeners are able to use fine-grained phonetic information in the speech signal to recognize spoken words.

However, these studies have mainly focused on coarticulatory properties or consonantal contrasts. They differ from the current dissertation as they have not explicitly considered the interaction between the acoustic realization of coarticulation and the need to maintain phonological contrast for similar phonetic realizations of vowels (e.g. coarticulatory vowel nasalization and contrastive vowel nasality). As CF has a phonological contrast between nasal and oral vowels (Léon, 1983; Martin, Beaudoin-Bégin, Goulet, & Roy, 2001) in addition to coarticulatory nasalization of oral vowels followed by nasal consonants (Desmeules-Trudel & Brunelle, 2018; Léon, 1983), it is particularly interesting to use CF as a case study for questions about the representation of phonetic detail and the interplay between coarticulation and contrast for lexical recognition.

In CF, the duration of nasalization, which corresponds to the timing of velum lowering and raising during the production of the vowel, is variable within and across categories: phonological nasal vowels and coarticulatory nasalized vowels can be nasalized for a longer or shorter period based on linguistic (e.g. phonotactic environment, prosodic position; Desmeules-Trudel & Brunelle, 2018), para-linguistic (e.g. geographic origin, register; Léon, 1983; Martin et al., 2001), and idiosyncratic factors (e.g. individual speakers; Arnaud, 2003, in Haut-Jura French). However, it is important to note that the duration of nasalization (i.e. velum lowering) is not the only articulatory cue to phonological vowel nasality in CF (Carignan, 2013; Delvaux, 2006), and that oral modifications to the vocal tract (e.g. tongue position) are also involved in the production of the oral-nasal contrast in French. The current project on the perception of the nasal contrast between coarticulatory and phonological vowels focuses on the duration of nasalization while keeping other phonetic characteristics constant (e.g. formant values, vowel duration). CF is therefore a language in which the processing of coarticulatory information can be compared to the processing of phonologically contrastive information, and in which within-category

variability in production could have an influence on the recognition of words that are distinguished based on the nasality/coarticulatory nasalization contrast.

Given the variability in the realization of the nasal-oral French contrast and the different patterns of coarticulation processing outlined above, the main research question of this work is:

- Is the detailed phonetic information related to vowel nasalization gradiently or categorically recognized by native listeners and second language listeners of CF in the context of spoken words?

The eye tracking experiments that are presented in this dissertation show that phonetic detail (i.e. variations in nasalization duration) is important during the recognition of spoken words. It also shows that time-dependent phonetic information within the segments, i.e. the moment when vowel nasalization starts within the vowel interval, modulates the interpretation of nasalized vowels in French, especially for L1 speakers and when stimuli contain *conflicting phonetic cues* (see Section 3.1.1.2). I also demonstrate that native listeners are more sensitive to small variations in the signal, resulting in gradient interpretation of vowel nasalization. On the other hand, L2 speakers display more categorical patterns of recognition overall, probably because they have had less experience in L2 French, and therefore less robust representations of the words' phonetic details, or are not able to access these fine-grained phonetic elements in their L2, given that processing demands are greater in the L2 than in the L1. Language transfer was also observed in L2 listeners, meaning that they sometimes tended to categorize L2 stimuli according to the L1 categories.

The implications of these findings are threefold. First, they enrich existing models of spoken word recognition such as TRACE (McClelland & Elman, 1986), which do not formulate clear predictions regarding the difference between L1 and L2 listeners. Second, this dissertation documents the patterns of phonetic integration of vowel nasalization within a language that has

contrastive nasal vowels. To my knowledge, this has never been done using eye tracking, a technique that enables one to dynamically track language processing (see Section 1.3). Third, I am able to assess the interference of a native language in L2 spoken word recognition, specifically by investigating individual listeners' characteristics such as proficiency and exposure.

The remainder of this chapter presents a brief overview of models of phonetic integration and recognition, a phonetic and phonological description of vowel nasalization in CF, and a brief description of the eye tracking method (i.e. Visual World Paradigm; Huettig, Rommers, & Meyer, 2011). The second and third chapters present literature on general speech perception mechanisms, as well as two eye tracking experiments that examine the processing of vowel nasalization in Canadian French with L1 listeners. In the fourth chapter, I review experimental and theoretical accounts of speech perception and spoken word recognition in a L2, focusing on models that were developed to account for naïve discrimination of foreign sounds as well as the acquisition of L2 sound categories. Two additional experiments with L2 listeners of CF (L1 English) are presented in Chapter 5. These experiments show how listeners process vowel nasalization in their L2, which is not a contrastive property in their L1. The last chapter of this dissertation returns to the theoretical questions of gradient perception, sound learning in a L2, and the phonetics-phonology interface in light of the obtained results.

1.1 PHONETIC INTEGRATION AND RECOGNITION

In this work, a speaker-listener's phonology is viewed as a dynamic processing system which integrates phonetic acoustic cues into abstract categories (i.e. phonemes), and this system is embedded into the process of spoken word recognition (SWR). Models of SWR such as TRACE

(McClelland & Elman, 1986), SHORTLIST (Norris, 1994) and PARSYN (Luce, Goldinger, Auer, & Vitevitch, 2000) assume a lower level of acoustic perception, an intermediate level of phonemic categorization, and a level in which lexical items are stored and activated. In these models, raw acoustic cues (i.e. features) activate phonemes, which activate lexical items while the phonetic input unfolds through time. When activation reaches a certain threshold, either at the phonemic or the lexical level, the phoneme or the word is recognized by the system and further grammatical and semantic information about the word can be accessed. Note that interactive models of SWR such as TRACE (i.e. models that assume both bottom-up and top-down influences during spoken word recognition) suggest that information from several levels (i.e. phonological and semantic) can be activated simultaneously, therefore that listeners do not need to “wait” for the word to be entirely pronounced before meaning is accessed. However, the purpose of this work is not to propose major modifications to existing models of spoken word recognition or claims about the inherent architecture of the models (levels of representation, activation/inhibition mechanisms, parallel vs. serial activation, etc.), but rather to use the proposed general SWR architecture as an interpretation framework for the obtained results.

One problem faced by the phonological and SWR systems is that phonetic cues do not all occur simultaneously during the production of phonemes. For example, voice onset time (VOT), following vowel length, pitch and first formant frequency are all phonetic cues linked to word-initial consonant voicing in English (Allen & Miller, 1999; Kessinger & Blumstein, 1998; McMurray et al., 2008). However, VOT occurs before the vocalic cues in the time-dependent signal. In perception, then, listeners may dynamically consider these various cues at different moments and continuously integrate them (McMurray et al., 2008), or alternatively listeners may buffer (Gow, 2003) the cues before making a decision on the phonological identity of the pronounced sound.

Similarly, authors have shown that a phonologically nasal vowel can be nasalized from onset onwards, or nasalization can start later in the duration (Delvaux, 2006; Desmeules-Trudel, 2015a; Desmeules-Trudel & Brunelle, 2018; Léon, 1983), which could potentially lead to the ambiguous realization of a phonological nasal ($\tilde{V}$) or coarticulated (VN) vowel. It is thus logical that listeners may consider this timing information when interpreting the phonological status of an early- or late-nasalized vowel. The vowel can be interpreted as a phonological nasal vowel or as a coarticulatory nasalized vowel, depending on when nasalization starts in the signal. The experiments presented in Chapters 3 and 5 were designed to test if variations in nasalization timing are considered, thus if the recognition of words that contain either $\tilde{V}$ s or VN is gradient or categorical, given the important within-category variability in their production.

1.2 VOWEL NASALIZATION IN (CANADIAN) FRENCH

In order to analyze the perception and interpretation of CF nasalized vowels, it is important to have a clear idea of the phonology of these vowels, and of their phonetic realization and variable characteristics. As mentioned above, CF uses vowel nasality as a phonologically contrastive property, meaning that words such as *paon* /pã/ ‘peacock’ and *pas* /pa/ ‘step’ constitute minimally different lexical items that are phonologically distinguished based on the nasality of the vowel. There are four contrastive nasal vowels in CF, namely /ẽ, ã, õ, œ/, as described by several authors (Côté, 2012; Martin, 2002; Martin et al., 2001).¹ These vowels contrast with oral vowels in all syllabic contexts (open or closed syllables – except [œ] for which

¹ In some systems of European French, /ẽ/-/œ/ have merged or the opposition between /ã/-/õ/ is neutralized, which suggests a simplification of the nasal vowel system. In CF, although /œ/ has the lowest functional load (Martin et al., 2001), all four nasal vowels are present and realized phonetically in minimal pairs.

no minimal pair is found in closed syllables, but for which [œ]-[œ̃] contrasts in the same closed syllabic context), as shown by the minimal pairs presented in Table 1.1.

Table 1.1

Minimal pairs of oral vowels and nasal vowels

	Oral – Opened σ	Nasal – Opened σ	Oral – Closed σ	Nasal – Closed σ
ε vs. ẽ	<i>fait</i> ‘fact’ [fɛ]	<i>fin</i> ‘fine’ [fɛ̃]	<i>faite</i> ‘done (f.s.)’ [fɛt]	<i>feinte</i> ‘feint’ [fɛ̃t]
a/ɑ vs. ɑ̃/ã	<i>pas</i> ‘step’ [pa]	<i>paon</i> ‘peacock’ [pɑ̃]	<i>patte</i> ‘paw’ [pat]	<i>pente</i> ‘slope’ [pɑ̃t]
o/ɔ vs. ɔ̃	<i>pot</i> ‘pot’ [po]	<i>pont</i> ‘bridge’ [pɔ̃]	<i>faute</i> ‘mistake’ [fot]	<i>fonte</i> ‘cast iron’ [fɔ̃t]
ə/œ/ø vs. œ̃	<i>je</i> ‘I (pro.)’ [ʒœ]	<i>(à) jeûn</i> ‘sober’ [(a)ʒœ̃]	<i>aveugle</i> ‘blind’ [avœgl]	<i>jungle</i> ‘jungle’ [ʒœ̃gl]

However, phonological nasal vowels have been repeatedly shown to display a high amount of variability, both in terms of their articulation and their acoustics, which is a key component of the current research. Sections 1.2.1 and 1.2.2 will review the literature on the articulatory and acoustic characteristics of CF and Québec French (QF) nasal vowels.

1.2.1 Articulation

French nasal vowels have different oral configurations than their oral counterparts (e.g. in European French see Carignan, 2014). Several articulatory features encode nasalization on the vowel, such as tongue position and jaw movement modifications to the articulation of nasal vowels when compared to their oral counterparts, in addition to velum lowering which enables air to go through the nasal cavity during phonation. In his detailed investigation of twelve speakers

of Northern Metropolitan (Paris) French², Carignan (2014) shows that for the [ɛ̃] vowel, the tongue tends to be lower and more retracted than for corresponding oral [ɛ]. Similarly for [ɑ̃], the tongue position is generally more retracted than for [a] (for 11/12 speakers), but tongue raising from [ɑ̃] towards a [ɔ̃]-like target (i.e. modification in the oral articulation of [ɑ̃]) is not consistent across participants. For the [o]-[ɔ̃] pair, speakers tend to produce the nasal vowel with a lower tongue position than [o]. Other articulatory differences, such as lip rounding and pharyngeal constriction, are also underlined. Although the general articulatory characteristics of CF nasal vowels are different from those of Northern Metropolitan French, this study convincingly shows that vowel nasality is not solely encoded by velum lowering during production. It also shows that variability is highly present in the realization of nasal vowels in French. This is crucial to the current research as I aim to verify if variability in production, more specifically the onset of nasalization on vowels, influences the perception and interpretation of vowel nasalization.

Studies on the oral articulation of nasal vowels in CF are not numerous and have not involved large samples of participants. However, Delvaux (2006) analyzed lip rounding, tongue ultrasound and airflow data from five native speakers of (Montreal) CF and found that, compared to European French, CF nasal vowels were generally less nasalized (i.e. nasal airflow rise onset occurred later for CF speakers), most vowels were diphthongized (especially /ɛ̃/ → [ɛ̃j] and /œ̃/ → [œ̃y]) resulting in fronting and rising of the vowels during their production, as well as increased lip rounding during the production of [ɔ̃]. These articulatory characteristics have a major influence on the acoustics of nasal vowels in CF, which in turn might have an influence on their perception and their representation in the phonological system.

² Note that the work presented here focuses on CF nasalized vowels which are different from European French vowels. However, as much of the work on the articulation of nasal and nasalized French vowels has been made on European varieties, it is pertinent to present these results to show that oral and nasal vowels do not necessarily have similar “oral” characteristics.

With regards to the duration of nasalization itself (velum lowering), which is the main phenomenon of interest to the current work, claims have been made about the late onset of nasalization on CF nasal vowels (often compared to European varieties of French). Delvaux (2006) and Desmeules-Trudel and Brunelle (2018) found a late rise of nasal airflow during the production of CF $\tilde{V}$ s, and some tokens for which nasal airflow was equal to zero after vowel onset, suggesting that velum lowering was not synchronous to the (preceding) consonant release: nasalization can be delayed in CF. Furthermore, Carignan (2013) and Desmeules-Trudel and Brunelle (2018) found that the peak of nasal airflow occurred after nasal vowel offset during the following stop consonant, yielding to the realization of a nasal consonantal appendix after the vowel, similar but perhaps less salient than the well-documented phenomenon of $\tilde{V}$ decompaction in Southern European French (Coquillon & Turcsan, 2012).

Although descriptions of the nasal appendix in CF are scarce, Desmeules-Trudel (2015a) and Desmeules-Trudel and Brunelle (2018) found that it is pervasive (i.e. it appears in more than 95% of $\tilde{V}$ s that are followed by an oral stop). Furthermore, according to Desmeules-Trudel and Brunelle (2018), it seems like the peak of nasal airflow of appendices is similar to that of nasal consonants, suggesting that the former display a significant amount of velum (see Figure 1.1, which reflects the “amount of nasalization” of nasal appendices and consonants). However, nasal appendices are short (i.e. 28 ms in open syllables and 50 ms in closed syllables, on average), and are the result of progressive coarticulation from a nasal vowel to the closure of a following oral stop (also in Brazilian Portuguese, see Desmeules-Trudel & Brunelle, 2018; Raposo de Medeiros 2011; Raposo de Medeiros, D’Imperio, & Espesser, 2008). It is thus expected that CF $\tilde{V}$ s are realized with a following nasal appendix in some contexts (e.g. when followed by an oral stop), but should still be considered as different from nasal consonants, at least in terms of duration.

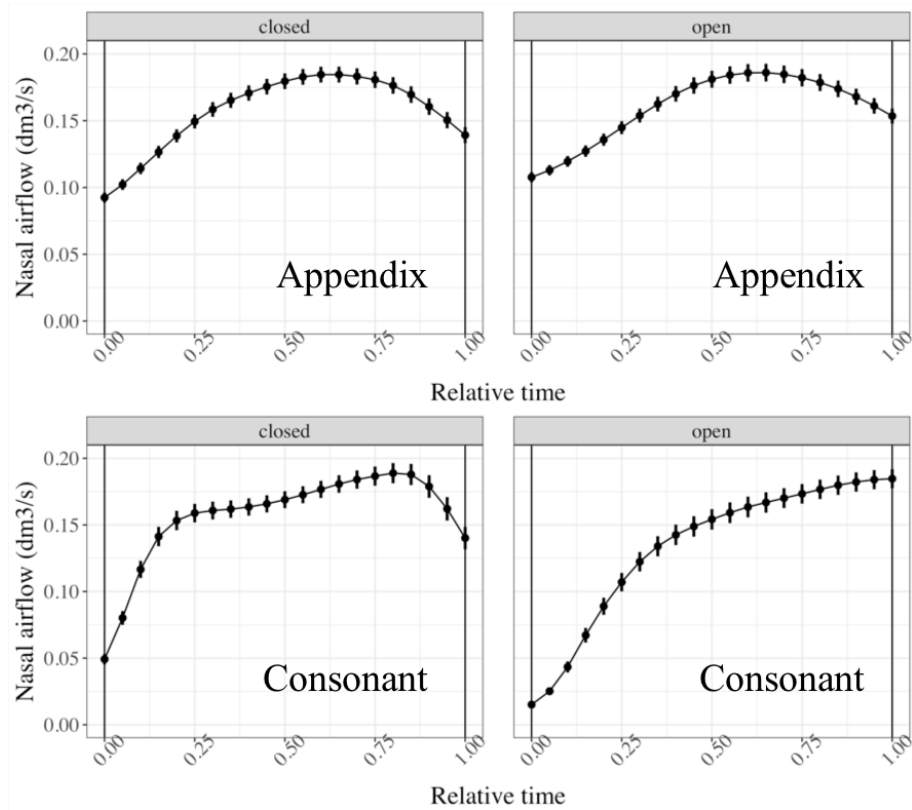


Figure 1.1. Nasal airflow patterns of nasal appendices (upper panels) compared to nasal consonants (lower panels) in closed (left) and open (right) syllables, from Desmeules-Trudel & Brunelle (2018)

1.2.2 Acoustic properties

Generally speaking, during the production of phonological nasal vowels and coarticulatory nasalized vowels, velum lowering causes a coupling of the oral and nasal tracts. This has a great influence on the acoustic spectrum (Figure 1.2) as it creates nasal formants and antiformants (Feng & Castelli, 1996; Fujimura, 1962; Kurowski & Blumstein, 1993; Maeda, 1993; Ohala & Ohala, 1993), flattens the lower part of the spectrum (Maeda, 1993), and causes a general loss of spectral energy (Delattre, 1965).

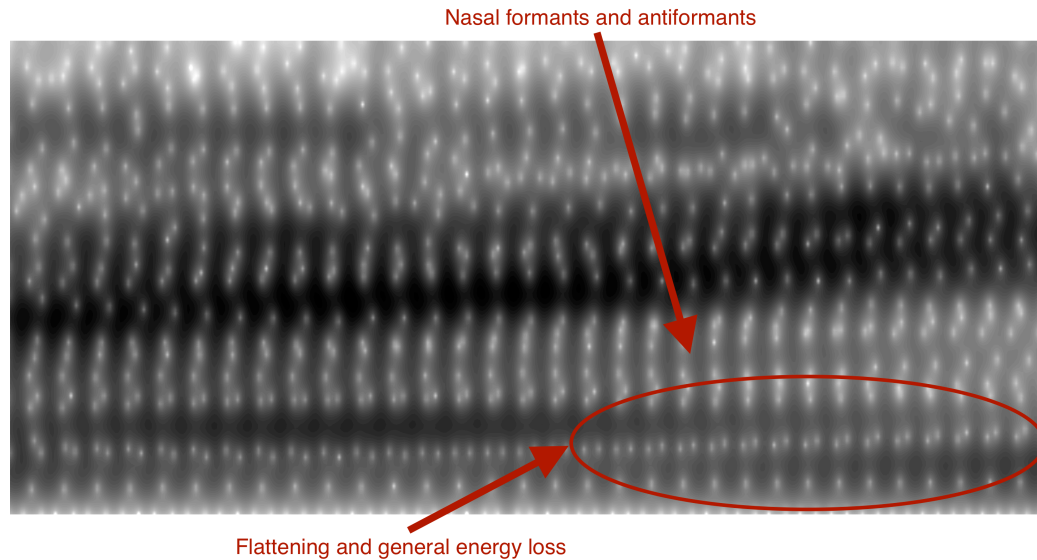


Figure 1.2. Spectrogram of the vowel [ẽ] (word *vin* [vẽ] ‘wine’) pronounced by a female speaker of CF, and acoustic correlates of vowel nasalization

Note that these acoustic effects vary depending on the quality of the vowel, as vowel height, anteriority/posteriority and lip-rounding also affect the spectrum. Idiosyncratic (i.e. individual) variability is also important, as anatomical differences such as the shape and composition of nasal sinuses seem to have an influence the acoustic shape of $\tilde{V}$ s (Dang & Honda, 1996; Dang, Honda, & Suzuki, 1994). Added to these effects are the language-specific acoustic characteristics of nasalized vowels, as well as additional oral configuration effects such as those described above which can cause modifications on the acoustic spectrum, and therefore influence the perception of vowel nasality in general or the quality of nasal vowels.

Acoustic studies of CF nasal vowels (Desmeules-Trudel, 2013; Léon, 1983; Martin et al., 2001) have also suggested that diphthongization occurs, as found in Delvaux’s (2006) articulatory study. Realizations such as /ã/ → [ã̃w̃], /ẽ/ → [ẽ̃j̃], /œ/ → [œ̃⁺] or [œ̃ỹ], and /õ/ → [õ̃w̃] are fairly frequent (Desmeules-Trudel, 2015a; Léon, 1983; Mielke, 2013; Martin et al., 2001;

Vajta, 2012). This factor is important for investigating differences between oral, nasalized and nasal vowels in CF as “oral” phonetic cues (i.e. cues that do not pertain directly to velum lowering) can potentially be used by listeners to differentiate $\tilde{V}$ s and VNs in perception or to recognize spoken words. However, most authors agree on the fact that diphthongization is variable in CF, therefore oral and $\tilde{V}$ s *can* be realized similarly with regards to their “oral” acoustic characteristics.

Overall, then, phonetic studies on the articulation and acoustics of vowel nasalization in (Canadian) French have underlined the variability of these vowels in production. As it has been shown that listeners do use fine-grained phonetic cues to recognize spoken words (see sections 1.1 and 2.2), it is likely that CF listeners are able to use variability in the signal during the production of nasalized vowels for perception and recognition purposes. The current experiments (Chapters 3 and 5) present evidence that variations in the duration of nasalization, which is one key aspect on which nasalized vowels are not realized consistently, do have an impact on the recognition of minimal pairs containing $\tilde{V}$ s and VNs (e.g. *pain* [pɛ̃] ‘bread’ – *peigne* [pɛŋ] ‘comb’), and that recognition can be gradient in native listeners of CF.

1.3 EYE TRACKING AND THE VISUAL WORLD PARADIGM

Several investigations of phonetic perception and phonological processing have used behavioural, offline techniques in order to tap into a listener’s phonological system. Traditional methods include discrimination tasks, in which two or more stimuli are presented one next to another in time, and listeners are asked to decide whether the two stimuli are the same or if they are different (AX task), or if one is similar to the first or second stimulus (ABX or AXB tasks; Liberman, Harris, Hoffman, & Griffith, 1957). Other variations of these discrimination tasks also

exist (4AIX – Pisoni, 1973; 2IFC – Schouten, Gerrits, & van Hessen, 2003), but most of them imply a categorical response from the listener: Same or different? Which member of the pair is similar to the target stimulus? Which pair of stimuli contains the *odd* one? For example, Gerrits and Schouten (2004) showed that categorical perception patterns emerged in simpler discrimination tasks (e.g. 2I2AFC) using a vowel continuum (from /u/ to /i/) with native Dutch speakers. In the 2I2AFC task, participants have to determine the order of the stimuli (AB vs. BA). However, Gerrits and Schouten (2004) did not find a categorical pattern of perception using a 4I2AFC task, which increases working memory load. In this task, participants are asked to determine in which position the odd stimulus occurs (e.g. second position for ABAA, third position for AABA). Furthermore, participants' judgements can be subjective as it can be difficult to infer which strategy they rely on to give their response. For example, in an AX task, some participants can base their response on phonological category membership and some on low-level acoustic differences (Werker & Tees, 1984). Categorical perception results thus seem to depend on the chosen task for investigation, and leaving more room for gradience and response variability could be beneficial for research on speech perception. This is one of the main reasons why the visual world paradigm (see below) was used for the experiments presented here, so as to avoid subjective judgements from the participants as much as possible by asking them to recognize meaningful lexical items, and to allow for gradient patterns of lexical recognition to emerge.

Another drawback of traditional discrimination tasks is the relative offline character of the responses given by the participant (e.g. button press, spoken repetition). However, eye tracking offers an online, dynamic measure of processing which does not require an overt response as eye movements are a reflex, and enables researchers to examine processing as time unfolds (Allopenna, Magnuson, & Tanenhaus, 1998; Cooper, 1974; McMurray et al., 2002; Salverda,

Brown, & Tanenhaus, 2011; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). These are major advantages for the use of eye tracking in language studies as speech is time-dependent.

One of the assumptions behind eye tracking is the *linking hypothesis* (Salverda et al., 2011) that participants tend to automatically direct their gaze on a visual scene towards what they are paying attention to (Huettig et al., 2011; Salverda et al., 2011; Salverda, Kleinschmidt, & Tanenhaus, 2014), given that attention is mediated by utterance meaning in the case of language studies. This enables researchers to tightly time-lock linguistic events in the speech signal (e.g. the onset of a word, the onset of a clause), among many possible manipulations, and investigate how listeners process language in real-time. For example, if two (or more) images are presented on a screen and an auditory word is presented simultaneously, a listener will direct their gaze towards the image that corresponds to the heard word, and recording and analysis of these eye movements can yield to conclusions on how spoken language is interpreted. During a typical trial, listeners see a number of images on a computer screen, usually a set containing at least a target and a competitor, and are instructed to perform a task using a mouse (e.g. “Place the book in the box”; Tanenhaus et al., 1995) or to listen-only to the stimuli (e.g. “Look at the book”; Johnson & Zamuner, 2010). Although it can be argued that this technique does not necessarily reflect *natural* language processing (see review by Salverda & Tanenhaus, 2017) or that the time it takes to plan an eye movement cannot allow researchers to investigate fine-grained properties in the speech signal (Salverda & Altmann, 2011), it is nowadays a relatively uncontroversial method of investigation for language processing studies (Salverda et al., 2011).

The particular eye tracking methodology of presenting a visual scene or a set of images simultaneously to auditory stimuli is referred to as the Visual World Paradigm (VWP; Huettig, et al., 2011). It has been extensively used in psycholinguistic research on speech sound processing (e.g. Beddor et al., 2013; Dahan et al., 2001b; Huettig et al., 2011; McMurray et al., 2008;

Zamuner et al., 2016), as well as other linguistics modules (e.g. Alamri, 2017; Andrews, Miller, & Rayner, 2004; Tanenhaus et al., 1995; van Rij, Hollebrandse, & Hendriks, 2016), and has enabled researchers to uncover many factors influencing language processing. As the current work focuses on the influence of vowel nasalization duration and signal variability on the recognition of spoken words and phonological processing, the VWP and associated tasks will be used as the main method in the experiments presented below.

2 Native language perception and word recognition

This chapter reviews some literature on speech perception and spoken word recognition in a native language in order to contextualize the experiments presented in Chapter 3. From categories and gradience in speech perception to spoken word recognition and lexical access, a number of factors influence how adults process and categorize the physical realization of phonemes to eventually recognize words in the speech stream.

2.1 CATEGORICAL PERCEPTION OF PHONOLOGICAL CONTRASTS

It has been shown at several occasions that distributional patterns (in production) *shape* the phonetic space in children and enable young language learners to develop phonological categories (Kuhl, 1992; 1994; Kuhl, et al., 2008; Pierrehumbert, 2003; Werker & Curtin, 2005), suggesting that the universal capacity to discriminate minimally different speech sounds becomes language-specific in the first few months of life. However, the question remains as to how sensitive listeners stay to variability in the signal for a given contrast after childhood. For example, one could ask if listeners categorically and consistently discriminate between two members of a contrast, or if they are able to perceive fine-grained phonetic information and use it for higher-level discrimination and identification purposes, such as spoken word recognition.

As a result of phonological category development, it has repeatedly been shown that listeners tend to categorize speech sounds from a continuum (e.g. VOT, formant values, vowel duration) onto discrete categories (e.g. voiced-voiceless, high-low vowels, long-short vowels). As an example, early work by Liberman et al. (1957) found that listeners tended to categorically label synthetic stimuli varying in the direction and magnitude on the second formant dimension,

replicating natural formant transitions, which indicate place of articulation of stop consonants in real speech. In their experiment, which also included an ABX discrimination task, category boundaries were sharp, suggesting that listeners were insensitive to variations within phonological /b, d, g/ categories with regards to second formant transitions. They were very accurate at distinguishing /b/ from /d/, for example, when the difference between the stimuli was physically similar to two instances of /b/, but crossed a phoneme boundary. This phenomenon, which most likely emerges from the phonemic category shaping during development, is known as categorical perception.

Several other studies have since then documented categorical perception in phonetics and language, but also in color perception and face recognition (Goldstone & Hendrickson, 2009; Harnad, 2003). It thus seems to be an inherent quality to human perception (as well as other species, see Kuhl, 1987), and recent definitions of categorical perception have focused on the fact that “perceived within-category differences are compressed and/or between-category differences are separated, relative to some baseline of comparison” (Harnad, 2003, p. 449). In the case of adult native speech perception, then, it can be hypothesized that the “baseline” which is referred to by Harnad, corresponds to the perceptual categories created by phonetic space shaping during language development, and that minimal acoustic differences (e.g. pitch or formant frequency, vowel duration, small variations in nasalization duration, small variations in VOT values, etc.) are compressed and ignored by higher-order systems such as the phonological system or the lexicon. In other words, the linguistic system filters out some of the non pertinent information, depending on the task demands, and this process results in categorical effects on (speech) perception.

It is not known, however, to what extent phonetic details are stored in memory, even if they might not be used in speech perception in general. Given categorical perception, in the case

of CF vowel nasalization, this should result in the absence of differences in participants' responses to slight variations in stimuli. Within the possible realizations of phonological nasal vowels, then, listeners should not make a difference between variable tokens, and similarly for VN sequences. The only notable differences should emerge across category boundaries, when the variations in nasalization timing are more extreme or cross the phonemic boundary between nasal and oral (coarticulated) vowels. However, it is also possible that listeners store some amount of phonetic details and are able to use them when task demands are greater, such as when they need to recognize spoken words. The experiments below will investigate this possibility.

2.2 GRADIENCE IN SPEECH SOUND PROCESSING

However, even though it seems likely that the general cognitive system does not process all the sensory information that is presented at every moment in our daily lives, investigation of categorical (speech) perception has focused on discrimination or identification tasks, which are intrinsically categorical. Therefore, it is possible that the ideas conveyed in research on speech perception, especially on categorical speech perception, may be biased by the types of tasks that have been used.

When using a method that leaves room for response gradience (see section 1.3), at least at the group level, research has found that gradient perception is possible, even for higher order processes that can be thought of as inherently categorical such as word recognition (McMurray et al., 2002). One notable example is McMurray et al. (2008) who showed that native English listeners were sensitive to small variations in VOT values (among other phonetic cues) on a continuum between voiced (e.g. /b/, VOT = 0 ms) and voiceless (e.g. /p/, VOT = 40 ms) consonants (see also McMurray et al., 2002). In their Visual World Paradigm (VWP) experiment,

listeners were instructed to listen to auditory stimuli composed of minimal pairs (e.g. *beach-peach*) that had been modified to form nine-step continua between 0 ms and 40 ms in VOT in 5 ms increments, and were asked to click on an image corresponding to the word they heard while their eye movements were recorded. Examining the listeners' fixation patterns (Figure 2.1), the authors found that unambiguous (0 ms and 40 ms) stimuli led to high proportions of fixations to the *correct* target (high proportions of fixations on the *beach* for 0 ms VOT, and high proportions of fixations on the *peach* for 40 ms VOT), but intermediate VOT values near the category boundary led listeners to display intermediate proportions of fixations values, as shown in Figure 2.1 from McMurray et al. (2008).

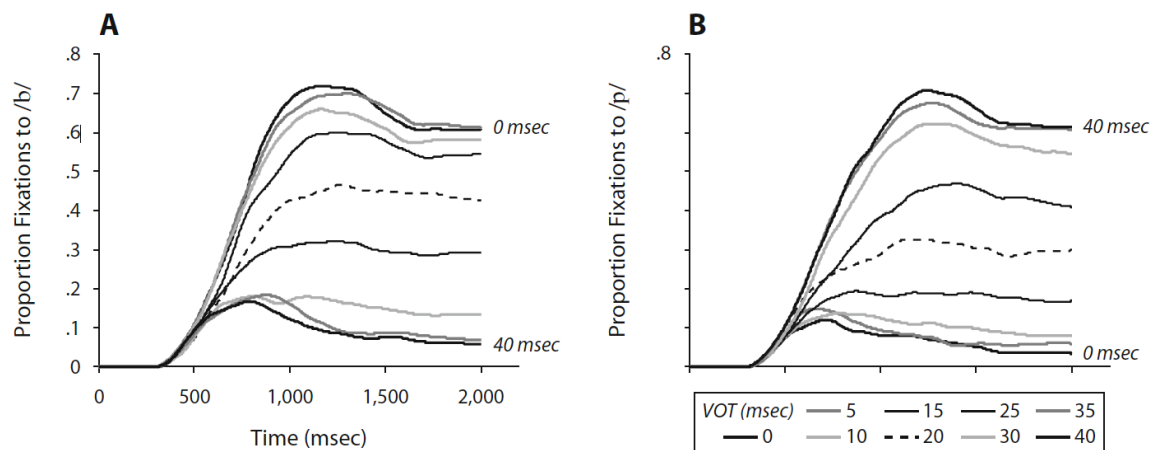


Figure 2.1. Participants' %fix patterns to the *beach* (panel A) or to the *peach* (panel B) on a 0-40 ms VOT continuum, from McMurray et al. (2008)

McMurray et al.'s (2002; 2008) studies are good examples of gradient patterns of lexical recognition based on small variations in the phonetic signal: listeners are able, at the group level and using an online measurement technique, to use these small acoustic differences in VOT for lexical purposes. In other words, the longer the VOT, the more /p/-like the consonant sounds,

inversely for shorter VOTs. Moreover, the boundary between /b/ and /p/ is not as sharp as what has been demonstrated in other studies using basic discrimination tasks (see Section 2.1). However, past studies on the integration of phonetic details has focused on consonants, and have not yet explored how fine-grained timing of *vocalic* cues are interpreted by listeners. I will thus focus on CF vowel nasalization timing, which is the main cue to the contrast between coarticulated (VN) and phonological ($\tilde{V}$) nasal vowels, in the remainder of this dissertation. In CF, if the gradient integration hypothesis is confirmed, small increments in the duration of nasalization on the stimuli are predicted to result in similar gradient patterns in word recognition. Vowels that are not nasalized should result in the interpretation of an oral vowel (followed by a nasal consonant) and vowels that are fully nasalized in the interpretation of a phonological nasal vowel. Intermediate values of nasalization duration (e.g. if a vowel is oral for its first half, and nasalized for its second half) should result in intermediate recognition patterns, as reflected by eye movement patterns.

2.2.1 *Other accounts of coarticulatory and gradient processing*

Although the literature on lexical recognition finds results in favour of the storage of fine-grained phonetic information, other accounts of coarticulation processing include theories of lexical storage for which coarticulatory features of phonological segments are underspecified (Lahiri & Marslen-Wilson, 1991). This implies that coarticulatory vowel nasalization, for example, would not be specified in the lexical representation of words containing VN sequences. Instead, knowledge of a phonetic rule of nasalization at the phonological level enables listeners to retrieve and anticipate the upcoming nasal consonant when hearing a nasalized vowel (Lahiri & Marslen-Wilson, 1991). It was thus expected that listeners would display categorical perceptual

patterns for coarticulated (VN) vs. phonological nasal ($\tilde{V}$) vowels in Bengali, which were found by Lahiri and Marslen-Wilson (1991). However, the authors did not consider nasalization timing variability in their analysis. The experiments presented in Chapters 3 and 5 will explicitly focus on signal variability by introducing nasalization timing differences in the stimuli, therefore allowing for comparisons between long-nasalized and short-nasalized vowels and evaluating how coarticulatory vs. phonological nasalization is represented in French in the context of spoken word recognition.

Other ideas have emerged from the literature in theoretical phonology concerning the possibility of having gradient representations of phonological contrasts. For example, in discussing the link between phonetics and phonology, Cohn (2006) gives the example of a sound change in progress for which the intermediate steps of progression cause ambiguity in the phonological status of the targetted contrast if a language community has not entirely adopted the change (actuation problem; Baker, Archangeli, & Mielke, 2011; Desmeules-Trudel, 2015b; Magloughlin, 2018). As sound changes are progressive through time, there might be continuity and gradience in the representation in the speech community. In this case, inter-speaker variability is expected to be greater before the sound change is complete.

Concerning contrast *well-formedness*, Cohn (2006) suggests that contrast can be viewed as gradient when considering phonotactics or functional load. For example, as it remains only in open syllables and before coronals, it is possible that the contrast between English [a] and [ɔ] is not as ‘categorical’ as other contrasts that occur in all syllabic positions. The same applies to the contrast between [θ] and [ð], which have a lower functional load than [t] and [d], for example. These issues are also thoroughly discussed, along with an experimental perceptual account, in Stevenson and Zamuner (2017) who show that listeners judge different French vowels as more similar and faster when they are in an allophonic relationship (i.e. [y]-[ʏ]) or when the contrast is

realized in some more specific context (restricted phonotactics; i.e. [o]-[ʊ]) than when they are in fully contrastive relationship (i.e. [a]-[ɔ]). This suggests gradience in the strength of phonological relationships across different segments. Although this vein of research will not be pursued in the current work, it is important to note that phonologists have considered the possibility of gradience (Hall & Hall, 2016) rather than the traditional approach of characterizing phonological contrast as strictly categorical.

2.3 PHONETICS AND WORD RECOGNITION

Gradience in speech perception can be observed through the interpretation of spoken words, as phonetic characteristics that enable listeners to recognize phonemes ultimately translate into how actual words are understood. According to McQueen and Cutler (2010, p. 489), “[w]ord recognition [is] at the heart of speech perception”, as the ultimate purpose of speech (perception) is to understand a message conveyed by words. Consequently, the cognitive processes underpinning lexical processing are highly dependent on speech perception skills and phoneme recognition. Psycholinguistic models of spoken word recognition (SWR) have focused a lot on the relationship between the phonetic elements of a word which do not have meaning by themselves, and actual lexical items which do have semantic content. Evidence from numerous studies has been found in favour of interactive models of SWR which allow for levels of representation to interact together in both bottom-up and top-down directions (e.g. TRACE; McClelland & Elman, 1986), as opposed to autonomous models for which the activation spreads in a bottom-up fashion only (e.g. SHORTLIST; Norris, 1994). Furthermore, the word recognition system deals with thousands of lexical items with a very limited set of phonemes. This underlines

the fact that listeners are highly efficient in retrieving words, and therefore in general language comprehension.

As mentioned earlier, speech perception interacts with higher-order representations. In addition to the word form, lexical representations contain a variety of information, such as frequency and semantic-grammatical properties. For example, in phoneme monitoring tasks, lexical effects of word/nonword and lexical frequency have been shown to influence performance and reaction times (Cutler, 1987; Rubin, Turvey, & van Gelder, 1976; Segui & Frauenfelder, 1986). In their study, Segui and Frauenfelder (1986) found that participants were faster at detecting phonemes when preceded by a prime word that was semantically related (i.e. to induce preactivation of the target word) than unrelated, an essentially lexical characteristic that has an influence on a “phonological” task. Other classical examples include the phonemic restoration effect (Warren, 1970) and the well-know Ganong effect (Ganong, 1980), which is when lexical context biases recognition of a (ambiguous) phoneme towards the meaning that is consistent with that context. These effects suggest some top-down influence from the lexicon onto speech perception, prelexical tasks. This has been taken as evidence in favour of interactive models of SWR such as TRACE (McClelland & Elman, 1986), although these results can also be explained by autonomous models of speech perception (Norris, McQueen, & Cutler, 2000) when considering the inherent architecture of specific models and attentional factors. The evidence presented here on the interaction between low-level acoustic information and lexical factors thus justifies the use of real words for investigating speech perception and phonological processing with listeners. It is thought to be a more realistic way to tap into listeners’ abstract representation systems than purely auditory tasks such as simple syllable discrimination or identification.

Overall, the studies presented in this chapter suggest that the well-known phenomenon of categorical perception (Goldstone & Hendrickson, 2009; Harnad, 2003; Liberman et al., 1957),

which emerges from category shaping during infancy (Kuhl, et al., 2008), might not be as robust as previously thought by psycholinguists, phoneticians and phonologists. Gradience in recognition patterns has been found using complex discrimination tasks (Gerrits & Schouten, 2004) or modern techniques such as eye tracking (Allopenna et al., 1998; McMurray et al., 2002; McMurray et al., 2008; McMurray et al., 2009). One common ground of the studies which have found gradient recognition patterns is that they focused on the influence of low-level acoustic information on SWR.

The following experiments will thus focus on variability in the phonetic realization of spoken words using eye movements as an indication of the way CF listeners integrate low-level phonetic details. One important addition to the body of literature is that vowel nasalization in CF is contrastive. Past research has not investigated phonetic variations of a vocalic contrast, to my knowledge. In the experiments presented in Chapter 3, triads of minimal words containing oral, contextually coarticulated (VN) and phonologically nasal ($\tilde{V}$) vowels were used to demonstrate the effects of phonetic variability on SWR. It is thus assumed that CF listeners demonstrated performance on a speech perception task in a more realistic way, and that the obtained results generate powerful yet plausible additions to existing models of SWR.

3 L1 perception of Canadian French vowel nasality and nasal coarticulation during spoken word recognition

This chapter presents experimental evidence that demonstrates that small variations in the duration of vowel nasalization impacts Canadian French (CF) native (L1) listeners' spoken word recognition performance. This was shown using words that contained phonological nasal vowels (e.g. *pain* [pɛ̃] 'bread'; C $\tilde{V}$) and oral vowels followed by a nasal consonant which can be nasalized (e.g. *peigne* [pɛɲ] 'comb'; CVN). This phonological nasality VS. coarticulatory nasalization (C $\tilde{V}$ ~CVN) contrast was chosen because the duration of nasalization on the vowel is variable in the production of CF within and across categories (Desmeules-Trudel & Brunelle, 2018; Léon, 1983), and because nasalization duration influences the recognition of spoken words in other languages (Beddor et al., 2013; Lahiri & Marslen-Wilson, 1991).

The experiments below were designed to test whether nasalization duration has an influence on the recognition of spoken words with L1 listeners of CF, and to test how fine-grained phonetic information and within-category variability is recognized and used during spoken word recognition. Some research has suggested that it is gradiently used for recognition (McMurray et al., 2002), while traditional views on the topic have suggested more categorical patterns of perception (Liberman et al., 1957). Both experiments used the eye tracking method to assess the predictions formulated below by each hypothesis. Recall that this is because eye movements are sensitive to gradient patterns of recognition and processing, while other more traditional techniques that yield categorical responses are not as likely to output response patterns that can demonstrate continuous integration of phonetic cues (Gerrits & Schouten, 2004). The main question of this research is repeated here:

- Is the detailed phonetic information related to vowel nasalization gradiently or categorically recognized by native listeners of CF in the context of spoken words?

A number of predictions can be formulated regarding this research question. Firstly, it is expected that variation in the duration of nasalization will have an influence on the recognition of spoken words containing oral vowels followed by nasal consonants (CVN) and on the recognition of phonological nasal vowels ($C\tilde{V}$), based on how this contrast is realized in production. For example, Desmeules-Trudel and Brunelle (2018) showed that phonological nasal vowels ($\tilde{V}$) are generally nasalized for 50% or more of their duration (as measured by a steep increasing slope in nasal airflow values). Consequently, vowels that are nasalized for a longer proportion of their duration (50% or more; long-nasalized) are expected to be recognized mostly as phonological nasal vowels. On the other hand, vowels followed by a nasal consonant (VN) in a closed syllable can be nasalized (i.e. coarticulated) for approximately 20% to 25% of their duration towards vowel offset, according to the data presented in Desmeules-Trudel and Brunelle (2018). Therefore, vowels that are nasalized for a shorter period of time (20% or less of their duration; short-nasalized) are expected to be mostly recognized as oral vowels (followed by nasal consonants).

Reaction times (RTs) were also recorded throughout the experiments, as they reflect listeners' processing load (Luce, 1986). Participants are expected to display shorter RTs when a vowel with long duration of nasalization (50% or more) is recognized as a phonological nasal ($\tilde{V}$) or when a short-nasalized vowel (20% or less of its duration) is recognized as coarticulated (VN). In both cases, the physical characteristics of the signal (measured from speech production) are congruent with the categorization by the listener. Inversely, vowels that are long-nasalized (50% or more of the duration) and recognized as coarticulated (VN; perhaps in smaller proportion) and short-nasalized vowels (20% or less of the duration) recognized as phonological nasal ($\tilde{V}$) should

yield longer RTs. In these cases, the physical realization of the word is incongruent with the listener's categorization. Furthermore, it is expected that stimuli that are ambiguous (i.e. that contain conflicting phonetic cues or missing phonetic cues) will display longer RTs in general.

Regarding the gradient or categorical recognition and use of phonetic cues, the hypotheses will be evaluated with the gradient or categorical patterns of chosen images eye movements. Recall that the gradient recognition hypothesis (McMurray et al., 2002) suggests that within-category variability has an influence on how listeners perceive and recognize spoken words. According to this hypothesis, variations in the duration of vowel nasalization are expected to yield gradient increase or decrease in proportions of fixations patterns to the target image. On the other hand, the categorical perception hypothesis (Harnad, 2003; Liberman et al., 1957) suggests that within-category variability is not considered by listeners, therefore that proportions of fixations to the target image will be more categorical and will not be influenced by variations in nasalization duration.

Concretely, patterns of predicted responses based on statistical models (chosen images and eye movements) will enable me to determine if word recognition was gradient or categorical. The used statistical procedure provides straightforward ways to predict data, which can then be plotted and visually assessed to confirm the categorical VS. gradient patterns of recognition. For example, if the GAMM-predicted data corresponds to a S-shape sigmoidal pattern with two plateaus (left panel in Figure 3.1), it can be inferred that recognition is categorical and that listeners are not sensitive to (or do not use) within-category variability. In this case, the change in category attribution is "steep" and conforms to the classical categorical-perception effect (Liberman et al., 1957). On the other hand, a linear pattern of recognition (right panel in Figure 3.1) suggests more sensitivity to within-category variability, since there should be a gradual shift in category attribution with the increase (or decrease) in nasalization duration. This conforms

more closely to the gradient-perception effect observed by McMurray et al. (2002). In addition to the visual inspection of GAMM predictions, statistical evaluation of the chosen images using the 50N %NASDUR value as a baseline will also enable me to verify if recognition was categorical or gradient. Since 50% nasalization duration on the vowel has been found to be at the boundary between $\tilde{V}$ s and VNs in production (Desmeules-Trudel & Brunelle, 2018), using this %NASDUR value as a baseline will enable me to assess differences across levels of the factor, and determine if participants responded in a categorical (e.g. no significant difference between 50N and 80N-100N) or gradient (e.g. significant difference between 50N and 80N-100N) manner.

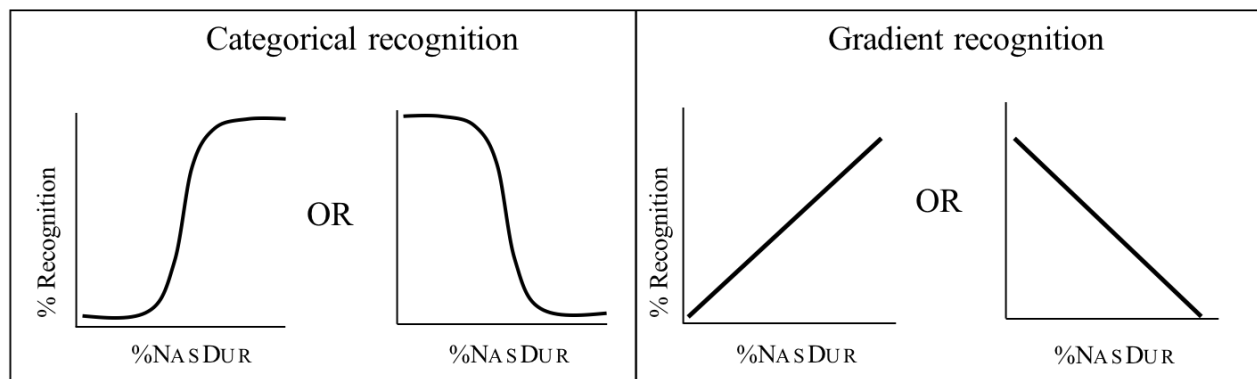


Figure 3.1. Possible predicted recognition patterns – Categorical (left) or gradient (right)

3.1 EXPERIMENT 1A

In order to assess the perception of phonetic details associated to vowel nasalization in CF, I used a Visual World Paradigm experiment. The stimuli in Experiment 1a contained variations in duration of nasalization as well as a short nasal consonant (i.e. analogous to the nasal appendix) at the end of the word. This means that stimuli contained at times conflicting phonetic cues, i.e. the presence of a nasalized vowel and a following nasal appendix in coda position. The choice was made to include a short consonant after the vowel to mimic the

production of phonological nasal vowels in connected speech, in addition to variations in nasalization duration. In a previous study on the aerodynamics of nasal vowels in Québécois French (Desmeules-Trudel, 2015a), 96% of the recorded vowels had a nasal appendix (i.e. a short, consonant-like, but not phonological, nasal segment found between the nasal vowel and following oral stop) based on productions of the vowels (i.e. 758/787 analyzed phonological nasal vowels had a nasal appendix, 39.8 ms duration, on average). Variations of nasalization duration were also included in the stimuli as this is the main focus of this work. The remainder of this section presents the methods (3.1.1), including more details on the stimuli (Section 3.1.1.2), results of the experiment (3.1.2), and a discussion of the results (3.1.3).

3.1.1 Methods

3.1.1.1 Participants

Twenty-four native listeners of CF were paid or received partial course credit at the University of Ottawa for their participation. Twenty-three were included in the current analyses (17 female, 6 male, between 18 and 36 years old, $\bar{X}$ = 23.3 years, SD = 5.3), and one had to be rejected from the analyses as they did not correctly fill out the consent. All listeners reported having normal hearing, normal or corrected-to-normal vision, and did not report any type or history of language, hearing or speech impairment. All participants completed a background language questionnaire (see Appendix 1), and self-reported knowing English as a second language at a moderate to high level of proficiency (see Figure 3.2), as well as other languages. Seven were French-English early bilinguals³, and sixteen learned English as an L2 at a later age.

³ To determine if listeners were early or late French-English bilinguals, a cut-off point of 30% or more exposure to English between the ages of 0 and five years old was chosen. Since research bilingualism level assessment based on exposure (Thordardottir, 2011) often establishes

Thirteen were born and raised in Ontario, nine in Québec, and one in New-Brunswick.

Importantly, no speaker had been exposed to another language containing contrastive nasal vowels in its phonological inventory.

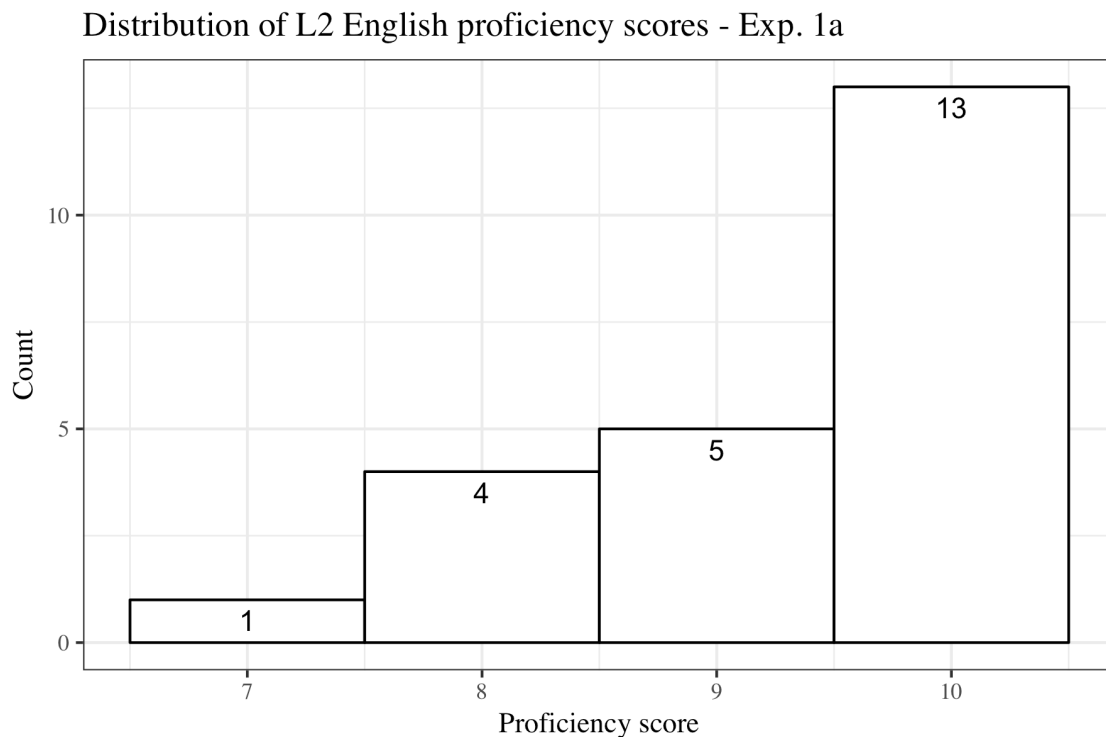


Figure 3.2. Distribution of self-reported L2 English proficiency scores⁴ out of 10 for Exp. 1a participants

its own criteria, I established the 30%-exposure threshold based on observation of the participants' self-reports. Between 0 and five years old, listeners had either between 0 and 10% exposure to English in one context (late bilinguals; $N=16$), or at least 40% (or more) exposure to English in at least one context (early bilinguals; $N=7$). The percentage of early exposure to English as a L2 is also thought to reflect the manner of acquisition (Sabourin, Leclerc, Lapierre, Burkholder, & Brien, 2016) within our groups: exposure to English as a L2 before age 5 tends to be more naturalistic (non-formal instruction), while participants who were exposed to English later tended to start learning it in school, in a more formal setting.

⁴ The L2 proficiency score corresponds to the sum of self-reported *spoken comprehension* and *spoken production* scores in English on the background questionnaire (see Appendix 1).

3.1.1.2 *Stimuli and experimental conditions*

Stimuli were nine triads of monosyllabic, picturable French words containing either a contrastive nasal vowel (C $\tilde{V}$), an oral vowel in a nasalization context (i.e. followed by a nasal consonant; CVN), or an oral vowel followed by an oral consonant (CVC); see Table 3.1. Note that the place of articulation of the oral consonant in CVC words did not always match the place of articulation of the nasal consonant in CVN words, which could have an influence on the results of the word-image association task. This will be discussed in Section 3.3. Three triads contained the front-mid vowel [ɛ- $\tilde{e}$], three contained low [a- $\tilde{a}$], and three contained back-rounded [ɔ- $\tilde{o}$]. Frequency of the nasal and coarticulated words could not be controlled due to constraints on word choice, but this factor was included in the statistical models for analysis. Raw word frequency values were extracted from the film subtitles subcorpus of the French language *Lexique* database (New, Brysbaert, Veronis, & Pallier, 2007) and log-transformed for statistical analysis (see Section 3.1.1.4). Using film subtitles as an estimation of word frequency better reflects the actual use of individual words in spoken language, and has been used in a variety of languages such as French (New et al., 2007), Dutch (Keuleers, Brysbaert, & New, 2010), Spanish (Cuetos, Glez-Nosti, Barbón, & Brysbaert, 2011), Chinese (Cai & Brysbaert, 2010), and British English (van Heuven, Mandera, Keuleers, & Brysbaert, 2014).

Table 3.1

Word list (experimental stimuli)

Contrastive nasal (C$\tilde{V}$) vowel	IPA	Frequency (log)	Coarticulated (CVN) vowel	IPA	Frequency (log)	Fully oral (CVC) vowel	IPA
<i>bain</i> ‘bathtub’	[b $\tilde{e}$]	1.703	<i>beigne</i> ‘donut’	[b ϵ n]	-0.319	<i>bec</i> ‘beak’	[b ϵ k]
<i>pain</i> ‘bread’	[p $\tilde{e}$]	1.798	<i>peigne</i> ‘comb’	[p ϵ n]	0.783	<i>pêche</i> ‘peach’	[p ϵ ʃ]
<i>vin</i> ‘wine’	[v $\tilde{e}$]	1.908	<i>veine</i> ‘vein’	[v ϵ n]	1.031	<i>verre</i> ‘glass’	[v ϵ ʁ]
<i>camp</i> ‘campsite’	[k $\tilde{a}$]	2.025	<i>canne</i> ‘cane’	[kan]	1.009	<i>cave</i> ‘basement’	[kav]
<i>dent</i> ‘tooth’	[d $\tilde{a}$]	1.123	<i>dame</i> ‘lady’	[dam]	1.937	<i>datte</i> ‘date (fruit)’	[dat]
<i>gant</i> ‘glove’	[g $\tilde{a}$]	0.994	<i>gamme</i> ‘musical scale’	[gam]	0.427	<i>gare</i> ‘bus/train station’	[gaʁ]
<i>pont</i> ‘bridge’	[p $\tilde{o}$]	1.703	<i>pomme</i> ‘apple’	[p ɔ m]	1.296	<i>porc</i> ‘pork’	[p ɔ ʁ]
<i>rond</i> ‘circle’	[ʁ $\tilde{o}$]	1.179	<i>rhum</i> ‘rum’	[ʁ ɔ m]	0.799	<i>robe</i> ‘dress’	[ʁ ɔ b]
<i>thon</i> ‘tuna’	[t $\tilde{o}$]	0.753	<i>tonne</i> ‘ton’	[t ɔ n]	0.651	<i>torse</i> ‘torso’	[t ɔ ʁs]

The word list was recorded by five native speakers of CF (2 female, 3 male, between 23 and 27 years old) from different regions in Canada: one from Southwestern Ontario (Windsor), one from Eastern Ontario (Ottawa), one from Québec (Montréal), one from New-Brunswick (Edmunston), and one from British Columbia (Victoria) in order to avoid listener habituation to one speaker during the test phase of the word recognition experiment. Four speakers were students in linguistics, and three speakers (two linguistics students and one professional) were simultaneous French-English bilinguals. The words were embedded into semantically appropriate sentences in order to avoid focus effects and were placed at the end of an intonational phrase in the carrier sentences for prosodic context consistency. Speakers were instructed to read the sentences twice in the most natural, but careful, way as possible, and the words in isolation were also recorded twice. A total of four tokens of each word was recorded from each speaker.

The recordings (words within sentences) were loaded into Praat version 6.0.08, and each word was hand-segmented and saved as a separate *.wav* file. All words were normalized for amplitude at 70 dB using a Praat script. C $\tilde{V}$ and CVN tokens were compared for each speaker individually, considering fundamental frequency and formant height, and most similar tokens were selected (total of nine word pairs) for cross-splicing. Similarity of the C $\tilde{V}$ -CVN pairs was considered due to potential differences in the articulation, and thus acoustic output, of each vowel quality (Carignan, 2014). Each speaker contributed one or two word pairs for the experimental stimuli. For each experimental stimulus, the consonant frame of a CVN token was kept by removing part of the oral vowel. Part of a C $\tilde{V}$ -token vowel was then pasted into the CVN frame (Figure 3.3). In the case of fully oral stimuli (see 0N in Table 3.2), the vowel from a CVC token replaced the CVN vowel entirely. For fully nasal stimuli (see 100N in Table 3), the C $\tilde{V}$ vowel replaced the CVN vowel. Vowel duration was adjusted when needed in order to match the original CVN value as much as possible (see below for more details on cross-splicing procedure).

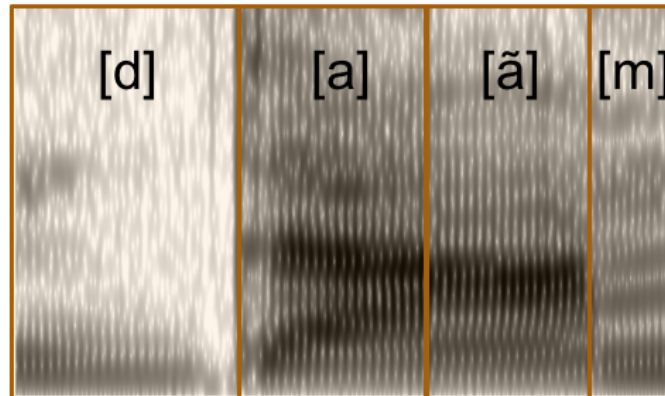


Figure 3.3. Spectrogram of *dame/dent* [dam]/[dã] ‘lady/tooth’ stimulus in the 50N %NASDUR – Exp. 1a

This yielded five experimental conditions (that were further converted in a continuous predictor for analysis, see section 3.1.1.4): fully oral vowels (0N), partially nasalized vowels (20N, 50N, 80N), and fully nasal vowels (100N). The conditions are illustrated in Table 3.2.

Table 3.2

Summary of the experimental conditions (%NASDUR) – Exp. 1a

		Vowel duration		
Condition	0N	Oral (100%)		50 ms nasal cons.
	20N	Oral (80%)	Nasal (20%)	
	50N	Oral (50%)	Nasal (50%)	
	80N	Oral (20%)	Nasal (80%)	
	100N	Nasal (100%)		

In the 0N condition, the oral vowel of a CVC token was pasted in the consonantal frame. For the 20N, 50N and 80N conditions, the duration of the oral vowel was calculated, and 20%, 50% or 80% of the vowel duration was cut from the original CVN vowel. The portion of the vowel that was removed from the original token was replaced by a part of the nasal vowel of the

matching C $\tilde{V}$ token at the zero crossing in order to avoid unwanted acoustic artifacts from the splicing such as clicks or noise in the signal. Amplitude peak reduplication was performed when necessary. In the 100N condition, the oral vowel of the original token was replaced by the full vowel of the matching C $\tilde{V}$ token. Final adjustments on fundamental frequency and amplitude were made, and the stimuli quality was verified by the experimenter (who is a native speaker of CF).⁵ Fifty ms of the final nasal consonant were kept in the stimuli in order to mimic the natural realization of nasal vowels' appendices in CF (Desmeules-Trudel, 2015a; Desmeules-Trudel & Brunelle, 2018). A total of 45 experimental stimuli was used in the experiment (5 stimuli for each word pair), and durational details about the stimuli are presented in Table 3.5.

Table 3.3

Auditory stimuli details

Stimulus⁶	Original duration	Final duration	Duration spliced	% spliced
Bain_0N	194.3	188.2	188.2	100
Bain_20N	194.3	184.9	42.1	22.77
Bain_50N	194.3	201	100.9	50.2
Bain_80N	194.3	201.2	159.3	79.17
Bain_100N	194.3	188.4	188.4	100
Camp_0N	189.4	195.7	195.7	100
Camp_20N	189.4	188.8	41.1	21.77
Camp_50N	189.4	180.5	96.5	53.46
Camp_80N	189.4	204.1	165.3	80.99
Camp_100N	189.4	193.4	193.4	100

⁵ Note that the stimuli were not independently validated with native speakers of CF. However, informal debriefing with each participant in Experiments 1 and 2 revealed that they found the stimuli natural-sounding, but sometimes ambiguous regarding their lexical identity.

⁶ Note that for all 0N stimuli, no portion of the phonological nasal vowel was spliced onto the matrix spoken word. Instead, the entire oral vowel from a CVC token was spliced onto the vowel from the CVN matrix word. Similarly, for 100N stimuli, the whole vowel from a CV token was pasted into the CVN matrix word. Consequently, only one F1 and one F2 value is given, at the vowel midpoint for 0N and 100N stimuli.

Table 3.3 (cont'd)

Stimulus	Original duration	Final duration	Duration spliced	% spliced
Dent_0N	177.8	176.7	176.7	100
Dent_20N	177.8	182.2	36.3	19.92
Dent_50N	177.8	172.6	84.7	49.07
Dent_80N	177.8	183.6	144	78.43
Dent_100N	177.8	168.2	168.2	100
Gant_0N	199.8	196.4	196.4	100
Gant_20N	199.8	197.6	42	21.26
Gant_50N	199.8	197	97.2	49.34
Gant_80N	199.8	207.5	163.9	78.99
Gant_100N	199.8	194.7	194.7	100
Pain_0N	192.5	196.7	196.7	100
Pain_20N	192.5	196.4	41	20.88
Pain_50N	192.5	192.6	96.9	50.31
Pain_80N	192.5	188.5	153.3	81.33
Pain_100N	192.5	192.5	192.5	100
Pont_0N	159.6	157.6	157.6	100
Pont_20N	159.6	159.5	35.7	22.38
Pont_50N	159.6	158.8	85.1	53.59
Pont_80N	159.6	159	122	76.73
Pont_100N	159.6	156.7	156.7	100
Thon_0N	176.6	149.1	149.1	100
Thon_20N	176.6	185.4	33.8	18.23
Thon_50N	176.6	171.2	93.7	54.73
Thon_80N	176.6	182.4	144.8	79.39
Thon_100N	176.6	144.2	144.2	100
Vin_0N	156.4	153.2	153.2	100
Vin_20N	156.4	157.9	31.1	19.7
Vin_50N	156.4	156.4	77.4	49.49
Vin_80N	156.4	156.1	130.6	83.66
Vin_100N	156.4	147.1	147.1	100

Visual stimuli corresponded to the words in the list and were taken from the International Picture Naming Project (IPNP) database (Székely et al., 2003), which contains black-and-white pictures of a variety of (English) nouns. Pictures of words that could not be found in the database (12 critical CŨ or CVN items, 7 non-critical CVC items, $N=19$) were hand-drawn by a

professional visual artist in the same style as the IPNP images, scanned and saved as *.jpeg* files. All visual stimuli were 5.6 X 5.6 cm (220 X 220 pixels), and images corresponding to the triads with one member containing a phonological nasal vowel (C $\tilde{V}$), a coarticulated nasalized vowel (CVN), and oral vowel (CVC) and a distractor were arranged together on a display. Images were embedded within approximately 6 X 6 cm (250 X 250 pixels) interest areas for data collection. For example, *pain* [p $\tilde{e}$], *peigne* [p ϵ ɲ], *pêche* [pɛʃ] and *pluie* [plɥi] (Figure 3.4) were presented on the screen. All experimental displays can be found in Appendix 2.

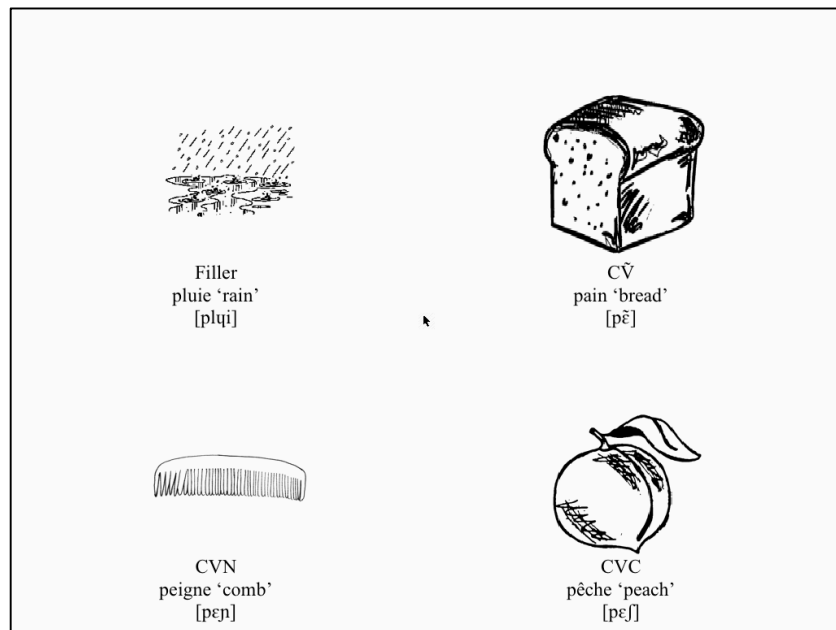


Figure 3.4. Example of one experimental display for the *pain* (C $\tilde{V}$; upper right) – *peigne* (CVN; lower left) pair along with *pêche* (CVC; lower right) and *pluie* (unrelated; upper left). Orthographic labels were not presented to the participants, but are included for illustration purposes.

3.1.1.3 Procedure

The experiment was programmed and presented in Experiment Builder (SR Research) version 1.10.63.⁷ After completing the consent form and background language questionnaire, participants were comfortably seated in front of the eye tracker (EyeLink 1000, SR Research) with their head stabilized on a chin rest at a distance of approximately 55 to 80 cm from the screen. The eye tracker was set to arm mount mode, with a 35 mm lens, sampling at 500 Hz. The procedure was explained to participants before the five-point calibration was performed. The calibration procedure was validated, and the maximum and average errors (in degrees of visual angle) were written down on the participant follow-up sheet. Errors were kept below 1° of visual angle to have better accuracy in eye movement measurements.

The familiarization phase started after the calibration. For each familiarization trial, participants saw an image on the screen with the corresponding unambiguous auditory word. Unspliced tokens of the critical C $\tilde{V}$ ($N=9$) and CVN words ($N=9$), CVC words ($N=9$), and monosyllabic filler words ($N=43$) were presented (total items, $N=70$). For each heard word, participants reported their degree of familiarity with the word by clicking on squares containing “PAS FAMILIER” (not familiar), “PEU FAMILIER” (little familiar), “ASSEZ FAMILIER” (fairly familiar), or “TRÈS FAMILIER” (very familiar). After completing the familiarization phase, participants were tested on the word-picture association task (i.e. experimental phase). In this phase, drift correction was performed before each individual trial: a black circle was presented in the center of the screen, and participants were instructed to look at the circle for the trial to begin. They then saw four images corresponding to the C $\tilde{V}$, CVN, CVC, and unrelated words. Images were on the display for 500 ms before an audio clip corresponding to an experimental (or filler) stimulus in one of the five experimental conditions (Table 3.2) was played. Participants had to

⁷ I want to thank Abdulrahman Alamri for sharing his experiment template.

click on the image that corresponded to the word they heard. Clicking on one of the images automatically ended the trial and changed the display to the next drift correction. At the drift correct screen, the experimenter indicated when the participant was looking at the drift correct black circle to move on to the next trial. No feedback was provided to the participants on the accuracy of their response, as some stimuli were inherently ambiguous. Image position randomly varied across trials to avoid preference (e.g. the upper left corner) for one image. Filler trials consisted of unambiguous word-image associations. There were 90 critical trials (i.e. each critical stimulus was presented twice), and 60 filler trials. The experiment lasted approximately 25 to 30 minutes. Participants were randomly assigned to one of four versions of the experiment in order to counterbalance the block presentation order.

The collected data included reaction times (RTs) between the onset of the stimulus vowel onset and the moment participants clicked on the corresponding image, fixations to the various images, and the images that participants clicked on. These measures were automatically compiled in Experiment Builder and extracted using SR Research DataViewer 1.11.1. Proportions of fixations to each image (i.e. fixations within the interest areas around the images) were calculated using a Python script provided by SR Research in 50 ms time bins.

Due to an error in programming, the clicked image data could not be compiled properly when some participants chose the image corresponding to the oral word (CVC) or distractor. The responses corresponding to the coarticulated image (CVN; e.g. *peigne*), the oral word (CVC; e.g. *pêche*), and the unrelated image (e.g. *pluie*) were all collapsed into the coarticulated (CVN) response category. In order to overcome this problem, the image that was fixated the most in the last 1000 ms of each trial was determined automatically using DataViewer, as it is thought to correspond to the image that was identified by the participants. Therefore, the CHOSEN IMAGE measure that will be presented below consists of the image that was fixated the longest in the last

1000 ms of each trial.⁸ The same measure will be presented for all experiments in the current dissertation.

3.1.1.4 *Analysis and variables*

Chosen images, RTs, and fixations to the C $\tilde{V}$ and CVN images (depending on chosen image, see below) were modeled using generalized additive mixed models (GAMMs; Wood, 2017). This statistical technique allows to analyze factorial and/or gradient predictors on potentially non-linear data, which is often the case in psycholinguistic data such as eye movements, therefore avoiding sphericity and sample-independence assumption violations encountered with traditional parametric analyses of variance (Max & Onghena, 1999). Furthermore, random effects structure (intercepts, slopes, and nonlinear smooth terms) can be added to the model, accounting for participant- or item-specific trends that could increase the risk of Type II errors. Also, auto-correlation of time series data is considered in this type of models (Wood, 2017), which is the case for time-dependent data such as eye tracking because one data point is correlated in time to the preceding point, and GAMMs can handle unbalanced data (i.e. missing data points). In the case of eye tracking data, missing data points are common as participants are free to fixate outside of the predetermined interest ideas during a trial, therefore creating lines in the data frame that are not filled with any fixations to the target, competitor, or other images on the display.

⁸ Combining participants from Experiment 1 and 2 for whom the corrected version of the experiment was used (i.e. their *CVC* and *filler* responses were correctly compiled, $N=12$), the CHOSEN IMAGE measure matches their mouse click responses 98.9% of cases. This allows me to be confident that the CHOSEN IMAGE measure captures the lexical choice that was made. There remains a margin of error with this measure, compared to using actual mouse click data, but as it was impossible to recover this information for a number of participants, the remainder of the chosen image analysis will focus on the image that was fixated the most in the last 1000 ms of each trial.

GAMMs have been used for investigating various types of linguistic questions, and have been used to analyze visual world eye movements (Porretta, Tucker, & Järvikivi, 2016; van Rij et al., 2016) as well as other time-dependent types of data in psycholinguistics and phonetics (e.g. ERP; Tremblay & Newman, 2015; nasal airflow; Desmeules-Trudel & Brunelle, 2018). The *mgcv* (Wood, 2017) version 1.8-16, *itsadug* (van Rij, Wieling, Baayen, & van Rijn, 2016) version 2.2, and *ggplot2* (Wickham, 2009) version 2.2.1 packages were used for the analysis and visualization of the data in the R environment (R Core Team, 2017) version 3.4.2.

For eye fixation GAMM modeling, the choice was made to use empirical-logit-transformed fixations to the C $\tilde{V}$ image (for C $\tilde{V}$ chosen image trials) and the CVN image (for CVN chosen image trials) as the dependent variable. *Empirical logits* correspond to an approximation of the log odds of fixating one given image on the screen and is an unbounded measure, which is suited to the inclusion of autocorrelation parameters within the models (Porretta, Kyröläinen, van Rij, & Järvikivi, 2018). An autocorrelation parameter cannot be included in binomial GAMMs. In other words, by using empirical logits (or a difference score based on empirical logits), correlation of one data point with the preceding one in the time series can be accounted for, but not using raw proportions of fixations due to the inherent 0-1 boundedness of the proportion measure. Empirical logits were calculated with the following equation:

$$(1) \log\left(\frac{y+0.5}{N-y+0.5}\right)$$

where y corresponds to the number of samples during which the “target” was fixated and N corresponds to the total number of samples within the time bin (i.e. given the 500 Hz sampling rate of the eye tracker and the 50-ms time bins, a maximum of 25 samples per bin was possible).

To fully assess (gradient) sensitivity to within-category variability, separate analyses were conducted depending on the participants' image choices: fixations to the C $\tilde{V}$ image were analyzed when participants chose the C $\tilde{V}$ image (C $\tilde{V}$ is assumed to be the target), and fixations to the CVN image were analyzed when participants chose the CVN image. Note that if all trials, regardless of image choice, were considered in the same analysis, gradient could emerge because of averaging across several trial types, and would not enable me to explicitly assess the influence of within-category variability on word recognition.

Several variables were considered in the statistical models for the GAMM analyses. These variables are described below and summarized in tables before the presentation of the results. The variables of interest are:

- PROPORTION OF NASALIZATION DURATION ON THE VOWEL (%NASDUR): numerical value between 0 and 100, corresponding to the proportion of the vowel duration that was nasalized (0, 20, 50, 80, 100 corresponding to 0N, 20N, 50N, 80N, and 100N). This is the main manipulation of this research, varying the duration of nasalization on the vowel. Note that this variable was mean-centered by subtracting the mean from each value in order to “increase numerical stability and remove any spuriously high correlation that may arise” (Tremblay & Newman, 2015, p. 128). For example, the 0N %NASDUR corresponded to -50 in the GAMMs, and the 50N %NASDUR corresponded to 0 in the GAMMs. For the chosen image analysis, this factor was discretized as five separate levels to assess the differences of each level to the baseline 50N %NASDUR condition.
- CHOSEN IMAGE (for RTs only): factor with two levels – Coarticulated image corresponding to the picture of the CVN word, and Nasal image corresponding to

the picture of the C $\tilde{V}$ word. This was measured based on the image that was fixated the most in the last 1000 ms of the trial. This variable was ordered and treatment-coded to assess the difference between the two levels of the predictor in the GAMMs.

- VOWEL QUALITY: factor with three levels – [a], [e], [o] roughly corresponding to the quality of the vowel for each word pair. Note that [ɛ] was noted [e] in the graph legends. This was specifically analyzed based on previous literature on differences in nasalization between vowel qualities (Carignan, 2014; Desmeules-Trudel & Brunelle, 2018). This variable was dummy-coded as it has more than two levels. Assessment of the differences across levels was based on [a].
- TIME: between 200 and 1300 ms after vowel onset, for the GAMM on eye tracking data only. This time window was chosen because 200 ms are necessary for eye movement planning (Fischer, 1992) after vowel onset. The end of the time window at 1300 ms was chosen based on the highest proportion of fixations value to the coarticulated image reached in the ON condition. Time was converted into a 0-1100 interval for the statistical model, where 0 is time-locked to the vowel onset, plus 200 ms for eye movement planning.

In the presentation of the results below (Section 3.1.2), focus will be on these preceding effects (%NASDUR, CHOSEN IMAGE for RT data, VOWEL QUALITY, and interactions through TIME for eye tracking data)⁹ in the main text. However, other variables were considered in the models as certain other factors pertaining to the lexicon (Dahan, Magnuson, & Tanenhaus, 2001a),

⁹ Note that time alone will not be included in the main text analysis, since it is not an effect that will enable me to draw conclusions on the use of variability for word recognition. The main effect of time is presented along with other supplementary factors in Appendix 3.

experimental design, and individual participants' characteristics can have an effect on the outcome of the experiment. While these variables are interesting and can provide insights into language processing, a full discussion of all of the variables distracts from the primary goals of the research. Rather than presenting these analyses here, illustrations of the significant results of the supplementary factors can be found in Appendix 3, and a discussion of the model results is included in 3.1.3. These factors were:

- FREQUENCY OF THE NASAL (C $\tilde{V}$) WORD: numerical value, extracted from the film subtitles subcorpus of the *Lexique* database (New et al., 2007), log-transformed and mean-centered (Tremblay & Newman, 2015).
- FREQUENCY OF THE COARTICULATED (CVN) WORD: numerical value, extracted from the film subtitles subcorpus of the *Lexique* database (New et al., 2007), log-transformed and mean-centered.
- TRIAL NUMBER: numerical value between 1 and 160, which reflects anticipation effects throughout the experiment.
- Self-reported PROFICIENCY in English as a second language: numerical value out of 10, including both productive and receptive proficiency in English.
- EXPOSURE to English as a second language before school age: binary factor that was coded as “1” (i.e. representing “yes”) if the participant had been exposed to English as a L2 30% of the time or more between the ages of 0 and 5 years old within the family, in kindergarden or with peers, and “0” (i.e. representing “no”) if they had been exposed to English less than 30% of the time in any context. This variable was ordered and treatment-coded to assess the difference between the two levels of the predictor in the GAMMs.

- PROVINCE OF ORIGIN: factor with three levels – Ontario, Québec, New-Brunswick, corresponding to the geographical origin of the participants. This variable was dummy-coded as it contains more than two levels.

Random effects structure was determined by maximum-likelihood chi-square tests (*compareML()* function from the *itsadug* package), comparing models with different random-effects structures (e.g. linear slopes by participants and items VS. non-linear smooth random terms by participants and items). For all final models, a backwards stepwise procedure starting with the full model, in which random effect structure had already been determined, was used: predictors and interactions from the full model were removed one at a time starting with the variables with the highest p -value. Predictors were kept in if one level combination of the factor or interaction had a p -value lower than 0.05. Note that p -values of parametric coefficients represent a significant difference of one level to the baseline, and that p -values of smooth terms indicate if it is different from 0. It indicates if a smooth term should be included in the model. Lower-order factors (numeric or parametric) and interactions were kept in to provide baselines for higher-order interactions even if they were not significant. For example, if the $\text{TIME}*\% \text{NASDUR}*\text{VOWEL QUALITY}$ was significant in the model, the individual TIME and $\% \text{NASDUR}$ smooth terms and the $\text{TIME}*\% \text{NASDUR}$ interaction were kept in the model, significant or not. Finally, the model was refitted using restricted maximum likelihood (REML). Note that this fitting procedure was used for all GAMM analyses throughout this dissertation, and has been used in past research to analyze similar data (e.g. Porretta et al., 2016).

3.1.2 Results

This section presents the results of Experiment 1a, focusing on the chosen images and reaction times (RTs) in Section 3.1.2.1. The eye tracking results are presented in Section 3.1.2.2.

3.1.2.1 Chosen images and reaction times

For the chosen image analysis, only the C $\tilde{V}$ and CVN chosen images were kept (1843/2070 trials), getting rid of the CVC and filler (i.e. incorrect) responses. In the upper panel of Figure 3.5, it is shown that participants chose the nasal image (C $\tilde{V}$; between 8.33% and 47.87%) less than the coarticulated image (CVN; between 52.13% and 91.67%) overall. The lower panel presents the GAMM predictions, and the output of the model is presented in Table 3.4. The GAMM analysis suggests that when compared to the baseline, the probability of choosing the C $\tilde{V}$ image in the 50N %NASDUR condition is significantly different (see Table 3.4) from all other levels of the factor. This suggests that the number of times the participants chose the C $\tilde{V}$ image gradually increased as the %NASDUR on the vowel increased.

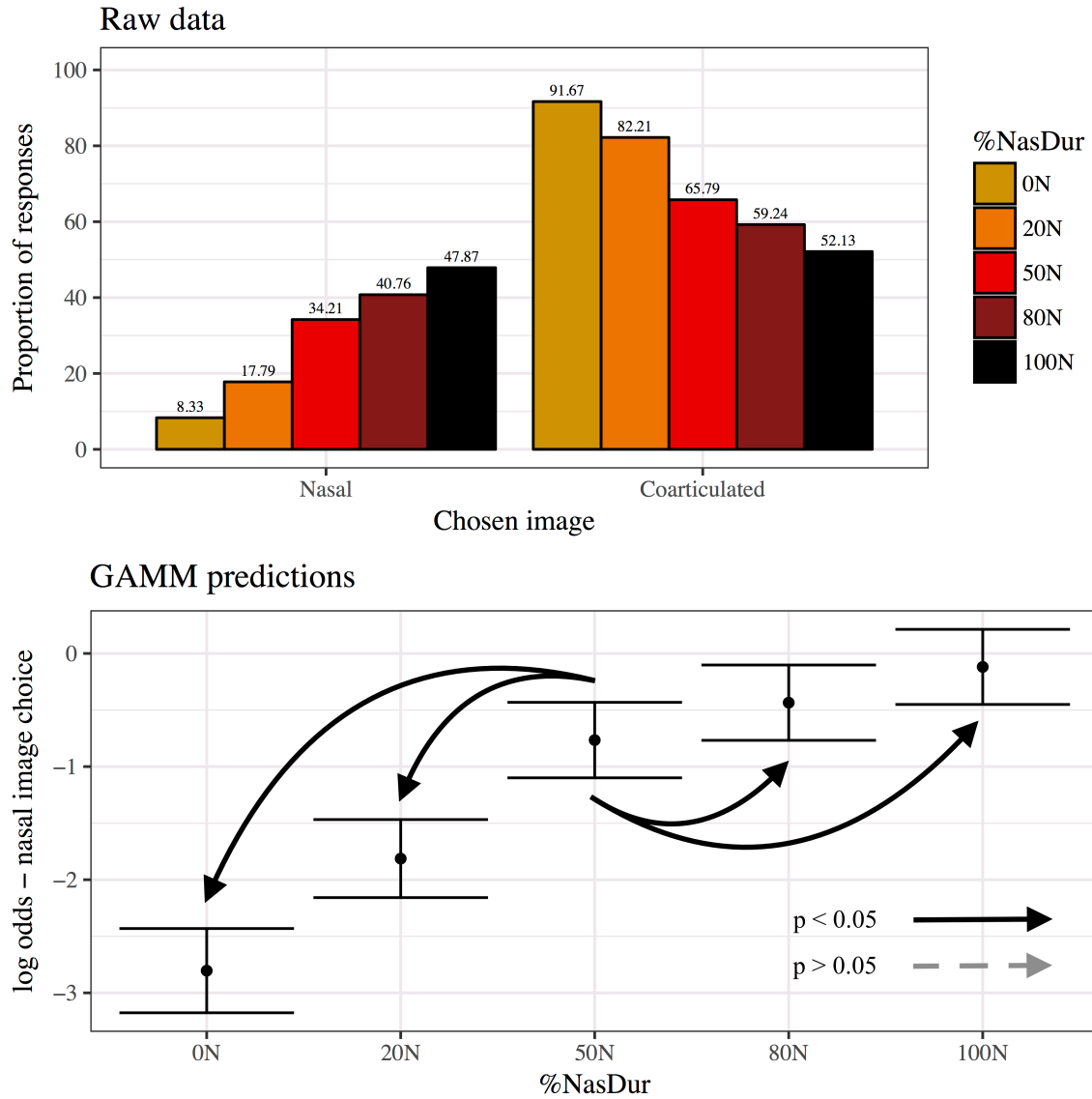


Figure 3.5. Proportions of image choices for nasal (C $\tilde{V}$) and coarticulated (CVN) images (upper panel) and GAMM predictions for image choice analysis (i.e. probability of C $\tilde{V}$ image choice by %NASDUR) – Exp. 1a

Table 3.4

GAMM results – Exp. 1a image choices

A. parametric coefficients	Estimate	Std. Error	t-value	p-value	
Intercept	-0.7658	0.3405	-2.2489	0.0245	*
%NasDur[0N]	-2.0379	0.2361	-8.6303	< 0.0001	***
%NasDur[20N]	-1.0473	0.1887	-5.5516	< 0.0001	***
%NasDur[80N]	0.3307	0.1669	1.9812	0.0476	*
%NasDur[100N]	0.6468	0.1653	3.9134	0.0001	***
B. smooth terms	edf	Ref.df	F-value	p-value	
Random_Participant	19.3283	22.0000	167.6144	< 0.0001	***
Random_Item	7.6396	8.0000	184.5792	< 0.0001	***

Concerning RTs (Figure 3.6), when participants chose the nasal (C $\tilde{V}$) image, there is an inverse relationship between RT and the duration of nasalization on the vowel. However, when participants fixated the coarticulated (CVN) image right panel), RT increased slightly when the vowel was more nasalized. When the vowel was not nasalized at all (0N %NASDUR), RT was slightly higher than when the vowel was slightly nasalized (20N %NASDUR).

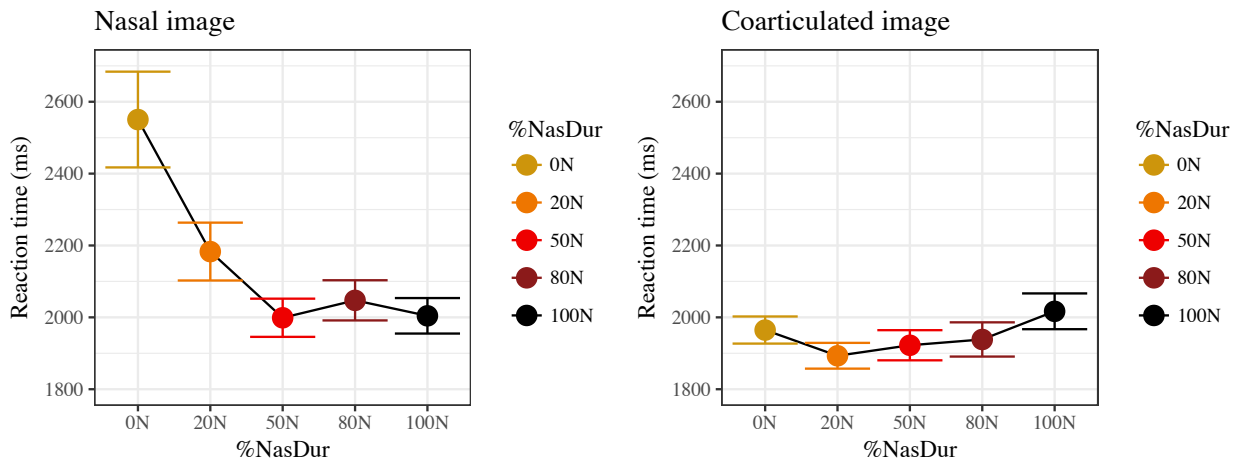


Figure 3.6. RT averages for nasal (C $\tilde{V}$) image choices by %NASDUR (left) and coarticulated (CVN) image choices (right) by %NASDUR – Exp. 1a

The remainder of this section presents the results of the GAMM analysis of RTs using the maximum likelihood (ML) method (for a similar model fitting procedure, see Porretta et al., 2016). The input variable of the model was the log-transformed RT. Predictors of interest and interactions are summarized in Table 3.5.

Table 3.5

Summary of the independent variables for the full RT GAMM – Exp. 1a

Parametric terms	Levels
CHOSEN IMAGE	Factor w/ 2 levels: Nasal, Coarticulated image
VOWEL QUALITY	Factor w/ 3 levels: [a], [e], [o]
PROVINCE OF ORIGIN	Factor w/ 3 levels: NB, ON, QC
EXPOSURE TO L2 BEFORE 5	Factor w/ 2 levels: True or False
Smooth terms	Levels
%NASDUR	Numerical (centered): -50 to 50
TRIAL NUMBER	Numerical: 1 to 160
LOG FREQUENCY (CVN)	Numerical: -1.0804 to 1.1756
LOG FREQUENCY (C $\tilde{V}$)	Numerical: -0.5229 to 0.7491
ENGLISH PROFICIENCY	Numerical: 7 to 10
Interactions	
%NASDUR*VOWEL QUALITY	
%NASDUR*FIXATED IA	
Random effects	
Smooth term:	
PARTICIPANTS*%NASDUR	
Smooth term:	
MINIMAL PAIR LABEL*%NASDUR	

Results of the final model are presented in Table 3.6 with CHOSEN IMAGE as a parametric factor, significant smooth terms for %NASDUR, TRIAL, and PROFICIENCY in English as well as a 2-way interaction between %NASDUR and CHOSEN IMAGE. Note that p -values lower than 0.05 here suggest that the smooth term is significantly different from zero, suggesting statistical difference between two or more points in the continuum (e.g. %NASDUR or TRIAL NUMBER). Presentation of the results will focus on the significant %NASDUR factor and

%NASDURXCHOSEN IMAGE interaction, as they are crucial to the research question. Illustrations of the additional significant factors (TRIAL NUMBER and ENGLISH PROFICIENCY) are found in Appendix 3.

Table 3.6

GAMM results – Exp. 1a RTs

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
Intercept	3.2771	0.0130	252.6785	< 0.0001
Chosen_image	0.0220	0.0078	2.8115	0.0050
B. smooth terms	edf	Ref.df	F-value	p-value
s(%NasDur)	1.0004	1.0006	0.1791	0.6721
s(Trial)	5.0848	6.1555	6.3470	< 0.0001
s(L2_prof)	1.0001	1.0002	11.2446	0.0008
s(%NasDur):Chosen_image	2.5727	2.8370	8.0725	< 0.0001
Random_Participant	19.1167	89.0000	2.4026	< 0.0001
Random_Item	26.1336	34.0000	7.0452	< 0.0001

Parametric coefficient estimates in Table 3.6 suggest that overall, when participants chose the coarticulated (CVN) image, RTs are significantly different from when they chose the nasal (C $\tilde{V}$) image (3.277 ± 0.022 , $t = 2.812$, $p > 0.01$). Note that nonsignificant s(%NASDUR) was kept in the model as a reference for the significant interaction. For this reason, it will not be illustrated below. Figure 3.7 present the data that is predicted by the model according to the significant %NASDURXCHOSEN IMAGE smooth term. To do so, the *get_fitted()* function from the *itsadug* package was used, outputting the fitted data and 95% confidence intervals (CIs) to the mean according to the model. These predicted data and CIs were then aggregated and plotted with the *ggplot2* package in R.

In Figure 3.7, the fitted RTs significantly depend on both the %NASDUR and the CHOSEN IMAGE. When participants chose the nasal image (C $\tilde{V}$), they responded slower to stimuli in the

0N %NASDUR (note that only 8.33% of the 0N trials were responded to as C $\tilde{V}$, i.e. 29/348 trials). In the 20N and 50N %NASDUR, they responded faster than in the 0N %NASDUR. In the 80N and 100N, participants had slightly longer RTs than in the 50N %NASDUR. This suggests a slight increase in RTs for vowels that are nasalized for 50% or more of their duration. On the other hand, when participants chose the coarticulated image (CVN), there was a slight increase in RT values overall between the 0N and the 100N %NASDUR.

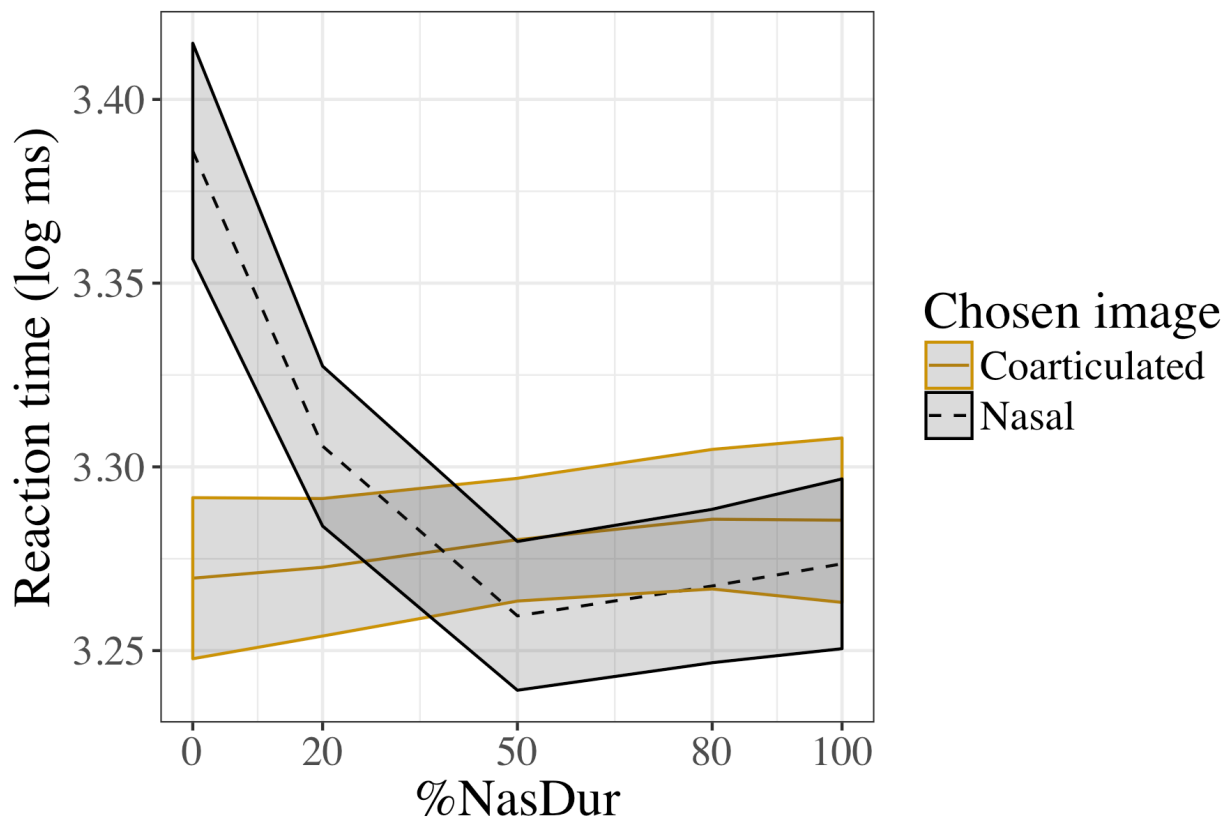


Figure 3.7. Predicted data by %NASDUR and CHOSEN IMAGE – Exp. 1a RTs

3.1.2.2 Eye movements

This subsection presents the eye tracking results, focusing on the proportions of fixations between 200 and 1300 ms after vowel onset. The beginning of the time window was chosen

based on the onset of the vowel and the 200-ms eye movement programming delay (Fischer, 1992). The end of the time window was determined considering the peak in proportion of fixations to the CVN target in the 0N condition when all chosen images are collapsed together, at 1300 ms. Patterns of fixations to the $C\tilde{V}$ (leftmost panels) and CVN images (rightmost panels) were plotted separately for when listeners chose the $C\tilde{V}$ image (upper panels) and when they chose the CVN image (lower panels). Note that participants chose the $C\tilde{V}$ image 555 times, and chose the CVN image 1288 times.

Let us recall that stimuli in this experiment were inherently ambiguous due to the presence of a short nasal appendix at the end of the vowel, in addition to actual nasalization of the vowel. This could have influenced the interpretation of the stimuli: two contradicting phonetic cues were presented, which can explain the bias towards choosing the CVN image. Furthermore, it seems like there is a positive relationship between the proportions of fixations to the $C\tilde{V}$ image and the duration of nasalization on the vowel throughout the time window (upper left panel in Figure 3.8). In other words, if the vowel was nasalized for a short period, participants looked less at the nasal ($C\tilde{V}$) image, especially when it was their final choice. This pattern suggests that the gradient phonetic details in the acoustic signal (i.e. duration of nasalization) is translated into the gradient interpretation of the vowels for word recognition purposes. The performed GAMM analysis below brings more insights into the question of gradience in word recognition based on nasalization duration.

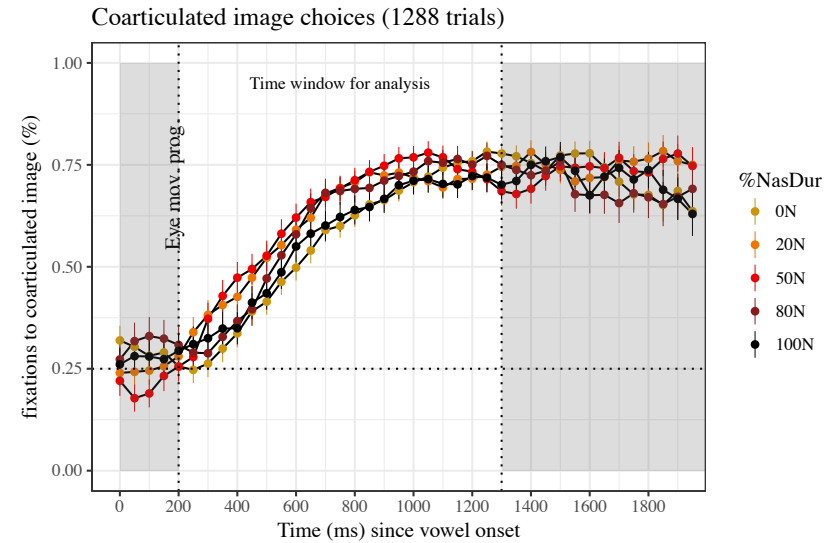
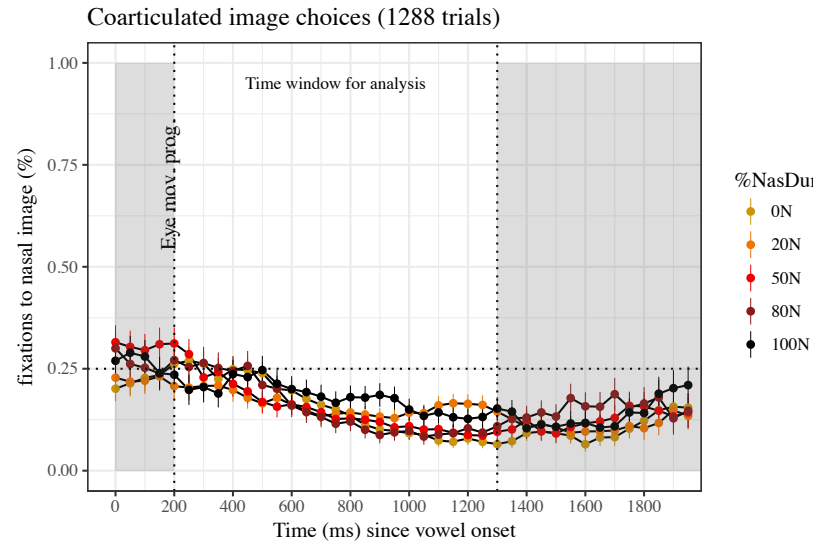
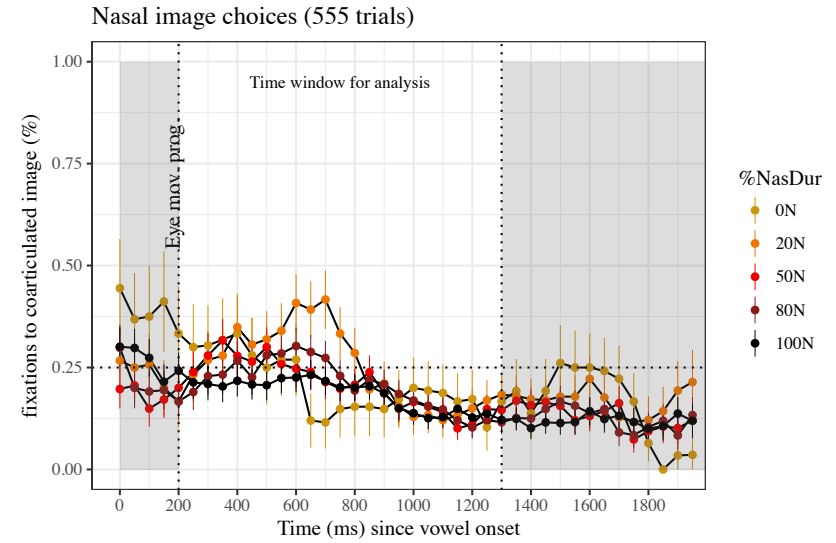
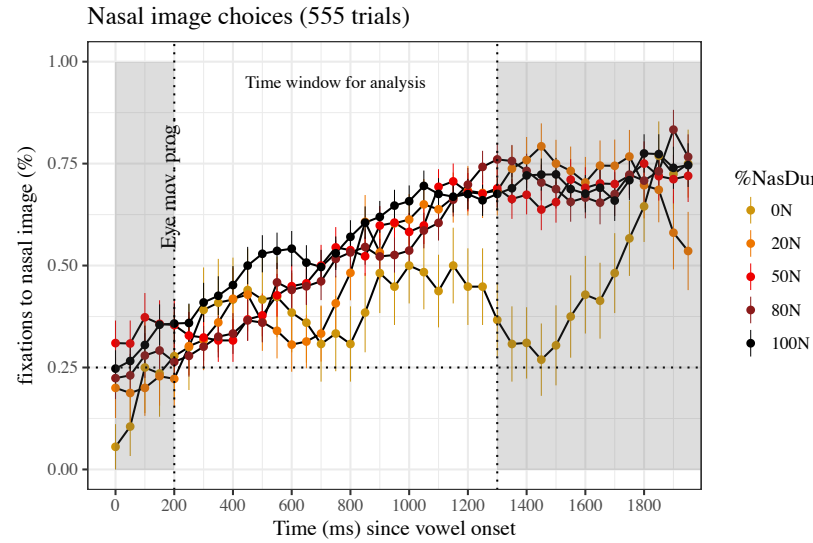


Figure 3.8. Proportions of fixations to the nasal ($C\tilde{V}$) image (leftmost) and coarticulated (CVN) image (rightmost), nasal ($C\tilde{V}$) choices (upper) and coarticulated (CVN) choices (lower) – Exp. 1a

The models that were fitted to the eye-tracking data used the maximum likelihood (ML) method (for a similar model fitting procedure, see Porretta et al., 2016) and a stepwise backwards fitting procedure described in Section 3.1.1.4. The input variable of the GAMM was the empirical logit-transformed fixations (Barr, 2008) to the nasal (C $\tilde{V}$) image in the first model, and to the coarticulated (CVN) image in the second model. The predictors of interest that were assessed in the statistical analysis are summarized in Table 3.7.

Table 3.7

Summary of the independent variables for the full GAMMs – Exp. 1a %fix

Parametric terms	Levels
CHOSEN IMAGE	Factor w/ 2 levels: Nasal (baseline), Coarticulated image
VOWEL QUALITY	Factor w/ 3 levels: [a], [e], [o]
PROVINCE OF ORIGIN	Factor w/ 3 levels: NB, ON, QC
EXPOSURE TO L2 BEFORE 5	Factor w/ 2 levels: True or False
Smooth terms	Levels
TIME	Numerical: 0 to 1100, separated in 50 ms time bins
%NASDUR	Numerical (centered): -50 to 50
TRIAL NUMBER	Numerical: 1 to 160
LOG FREQUENCY (CVN)	Numerical: -1.0804 to 1.1756
LOG FREQUENCY (C $\tilde{V}$)	Numerical: -0.5229 to 0.7491
ENGLISH PROFICIENCY	Numerical: 7 to 10
Interactions	
TIME* FIXATED IA	
TIME* VOWEL QUALITY	
TIME* PROVINCE OF ORIGIN	
TIME* EXPOSURE TO L2 BEFORE 5	
TIME*%NASDUR	
TIME* TRIAL NUMBER	
TIME* LOG FREQUENCY (COART.)	
TIME* LOG FREQUENCY (NASAL)	
TIME* ENGLISH PROFICIENCY	
TIME*%NASDUR *VOWEL QUALITY	
TIME*CONDITION*FIXATED IA	
Random effects	
Smooth term: PARTICIPANTS*%NASDUR	
Smooth term:	
MINIMAL PAIR LABEL*%NASDUR	

The first GAMM modeled (empirical-logit-transformed) proportions of fixations to the C $\tilde{V}$ image when participants chose the C $\tilde{V}$ image (i.e. 555 trials). The output of that model is presented in Table 3.8. Let us recall that the TIME factor only includes the window of analysis, taking into account the eye movement planning delay (0 ms actually represents the beginning of the analysis window, 200 ms after vowel onset). Presentation of the results will focus on the significant %NASDUR factor, and the TIMEX%NASDUR and TIMEXCHOSEN IMAGEXVOWEL QUALITY interactions. Illustration of the additional significant smooths (TIME, FREQUENCY (NASAL) and interaction between TIME and TRIAL) is found in Appendix 3. Each smooth term is detailed and illustrated below (Figures 3.9 to 3.11). Auto-correlation AR1 value, corresponding to the average correlation between one data point and the preceding one across the data, was empirically set to 0.836 based on the data.

Table 3.8

GAMM results – Exp. 1a %fix to C $\tilde{V}$, C $\tilde{V}$ choices

A. parametric coefficients	Estimate	Std. Error	t-value	p-value	
Intercept	0.2667	0.2968	0.8987	0.3688	
Vowel_[e]	0.5885	0.4460	1.3194	0.1871	
Vowel_[o]	-0.6321	0.3732	-1.6934	0.0904	
B. smooth terms	edf	Ref.df	F-value	p-value	
s(Time)	1.0049	1.0064	19.9519	< 0.0001	***
s(Trial)	2.6585	3.2926	1.5851	0.2081	
s(%NasDur)	1.0088	1.0172	4.8969	0.0270	*
s(Log_freqNasal)	1.0795	1.1176	13.3373	0.0002	***
s(Time:Trial)	12.4921	14.5873	4.9111	< 0.0001	***
s(Time:%NasDur)	6.3832	7.3994	4.0197	0.0001	***
s(Time:%NasDur):Vowel_[a]	5.4161	7.1572	0.9982	0.4009	
s(Time:%NasDur):Vowel_[e]	9.8288	12.4236	2.8956	0.0004	***
s(Time:%NasDur):Vowel_[o]	10.3510	12.4887	3.2425	0.0001	***
Random_Participant	103.4380	206.0000	1.3309	< 0.0001	***
Random_Item	44.2282	77.0000	1.4748	< 0.0001	***

The VOWEL QUALITY parametric factors was kept in the terms of the model to include an intercept for the significant smooth terms in B, and the non-significant main effect of TRIAL was also kept in the model to include a baseline for the significant TIME*TRIAL interaction. The following figures present the data that is predicted by the model for each significant smooth term of interest. To do so, the *get_fitted()* function from the *itsadug* package was used, outputting the fitted data and 95% confidence intervals (CIs) to the mean according to the model. These predicted data and CIs were then aggregated and plotted with the *ggplot2* package in R.

One of the main predictors of interest is the %NASDUR, which reflects the duration of vowel nasalization. Figure 3.9 shows that longer %NASDUR yielded more fixations to the nasal (C $\tilde{V}$) image. There is a relatively constant increase in overall fixations with the increase of %NASDUR (except between 0N and 20N nasalization, which represents a small number of trials, see Figure 3.8). This is expected based on the realization of vowel nasalization in CF: nasal (C $\tilde{V}$) vowels are nasalized for a longer period (roughly the 50N or more) than coarticulated (CVN) vowels (Desmeules-Trudel, 2015a). The observed pattern of gradual increase is partly consistent with the gradient recognition hypothesis.

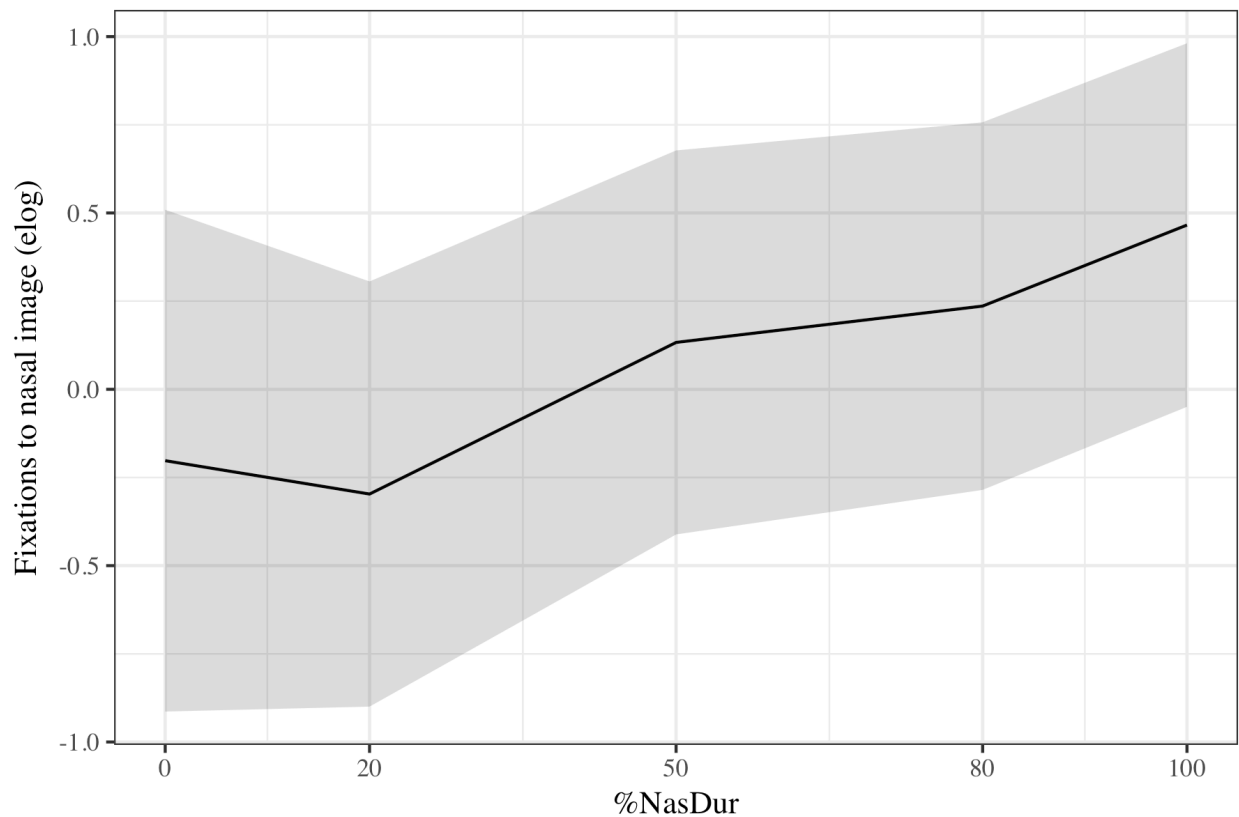


Figure 3.9. Predicted data by %NASDUR – Exp. 1a elog %fix to $\tilde{C}\tilde{V}$, $\tilde{C}\tilde{V}$ choices

Figure 3.10 shows the effect of the %NASDUR on the time course of fixations to the nasal image. As expected, participants fixated on the $\tilde{C}\tilde{V}$ image more through TIME when it was entirely nasalized (i.e. black broken line in Figure 3.10). This is not surprising as the heard vowel in the 100N %NASDUR condition canonically corresponds to the participants' expectations. There is also a decrease in proportions of fixations to the $\tilde{C}\tilde{V}$ image through TIME in the 0N %NASDUR. However, let us recall that this %NASDUR condition only includes 8.3% of 0N trials (i.e. 29/348 trials), which is a small number of trials to draw a firm conclusion on the participants' behaviour in this condition when they chose the $\tilde{C}\tilde{V}$ image.

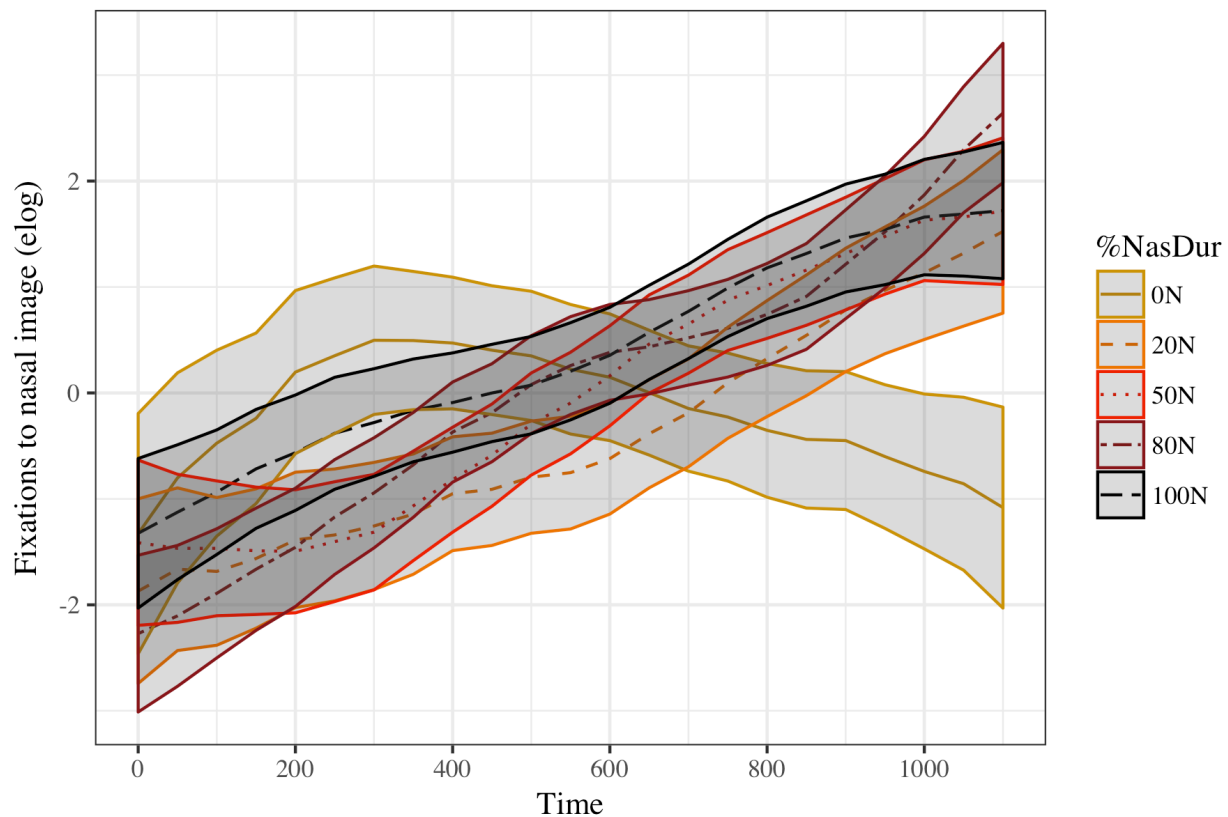


Figure 3.10. Predicted data by CHOSEN IMAGE through TIME – Exp. 1a elog %fix to $C\tilde{V}$, $C\tilde{V}$ choices

Shown in Figure 3.11 is the 3-way interaction between TIME, VOWEL QUALITY and %NASDUR. In this figure, different vowel qualities have slightly different patterns of fixations to the nasal ($C\tilde{V}$) image through time. Overall, the 0N %NASDUR behaves differently from the other %NASDUR values, except for [e] (middle panel) in which all conditions seem to overlap, to some extent. There is also some variation across CONDITIONS and VOWEL QUALITIES, e.g. the 50N %NASDUR for the vowel [e] dramatically decreases after 500 ms, as opposed to the other vowels. This may suggest that the ambiguous vowel [e], when nasalized for 50% of its duration, is

eventually interpreted as a CVN, while [e] that is nasalized for 80% or more of its duration tends to be interpreted as a phonological nasal vowel. Overall, the displayed pattern is complex.

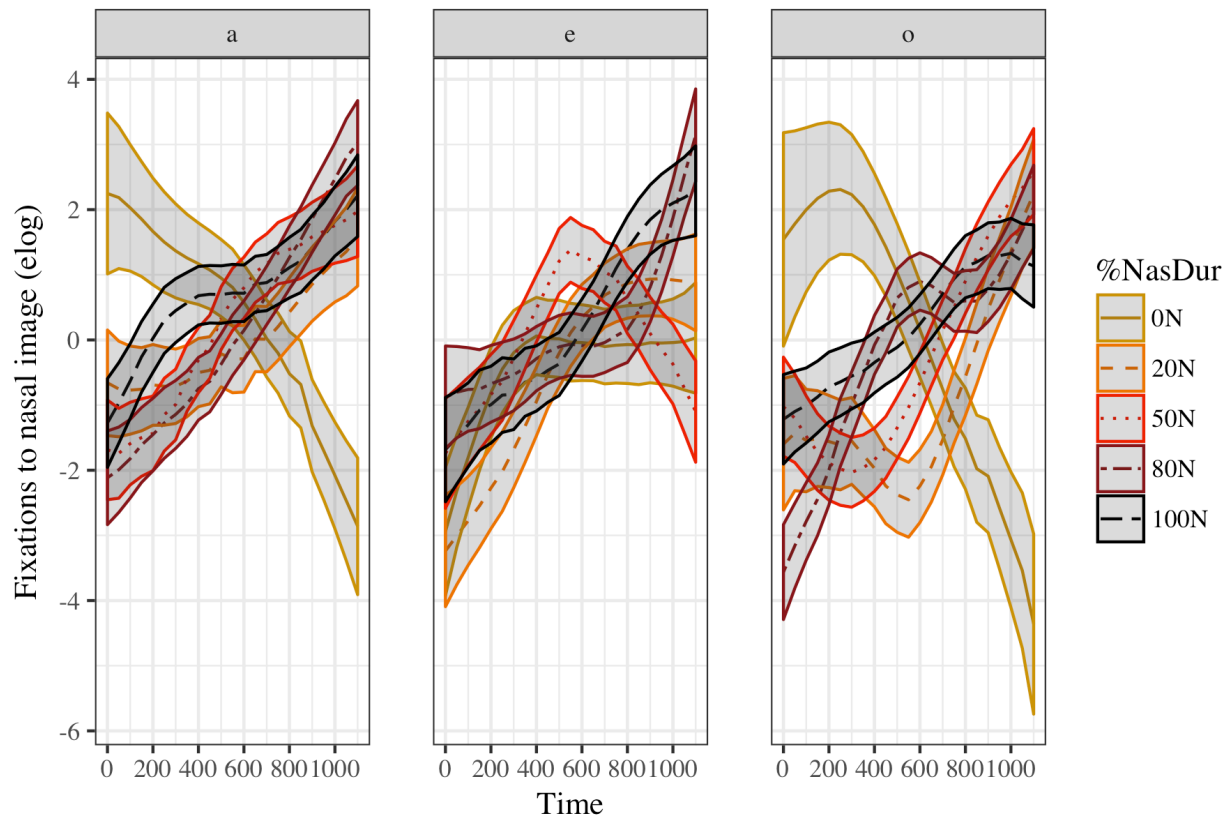


Figure 3.11. Predicted data by %NASDUR and VOWEL QUALITY through TIME – Exp. 1a elog %fix to C $\tilde{V}$, C $\tilde{V}$ choices

The second GAMM modeled (empirical-logit-transformed) proportions of fixations to the CVN image when participants chose the CVN image (i.e. 1288 trials). The output of that model is presented in Table 3.9. Presentation of the results will focus on the significant %NASDUR factor, and the TIMEXCHOSEN IMAGEXVOWEL QUALITY interactions. Illustration of the additional significant smooths (TIME, TRIAL, and interactions between TIME and PROVINCE, TRIAL, AND FREQUENCY (NASAL)) is found in Appendix 3. The smooth terms are detailed and illustrated

below (Figures 3.12 and 3.13). Auto-correlation AR1 value, corresponding to the average correlation between one data point and the preceding one across the data, was empirically set to 0.843 based on the data.

Table 3.9

GAMM results – Exp. 1a %fix to CVN, CVN choices

A. parametric coefficients	Estimate	Std. Error	t-value	p-value	
Intercept	-0.0979	0.8748	-0.1119	0.9109	
Vowel_[e]	0.3923	0.4942	0.7938	0.4273	
Vowel_[o]	0.2549	0.4555	0.5596	0.5757	
Province_ON	0.5775	0.8390	0.6883	0.4913	
ProvinceQC	0.9510	0.8458	1.1245	0.2608	
B. smooth terms	edf	Ref.df	F-value	p-value	
s(Time)	3.3543	3.7505	2.9090	0.0251	*
s(Trial)	6.0603	7.2124	5.2662	< 0.0001	***
s(%NasDur)	2.1260	2.4955	3.7972	0.0218	*
s(Freq_nasal)	1.2146	1.2442	1.2720	0.2071	
s(Time:Province_NB)	5.4365	6.1620	2.9363	0.0069	**
s(Time:Province_ON)	1.0229	1.0280	1.3092	0.2545	
s(Time:Province_QC)	2.9942	3.4079	1.4689	0.2311	
s(Time:Trial)	7.9970	10.5256	2.3757	0.0071	**
s(Time:%NasDur)	1.0360	1.0486	0.5115	0.4845	
s(Time:Freq_nasal)	2.9218	3.1039	4.7549	0.0025	**
s(Time:%NasDur):Vowel_[a]	10.1681	12.9962	2.0000	0.0168	*
s(Time:%NasDur):Vowel_[e]	10.2197	13.1441	2.7754	0.0005	***
s(Time:%NasDur):Vowel_[o]	8.7488	11.7189	3.0646	0.0002	***
Random_Participant	125.5171	205.0000	2.2410	< 0.0001	***
Random_Item	50.9505	77.0000	3.3605	< 0.0001	***

The VOWEL QUALITY and PROVINCE parametric factors were kept in the terms of the model to include an intercept for the significant smooth terms in B, and the non-significant main effects and interaction of FREQUENCY (NASAL) and TIME*%NASDUR were also kept in the model to include a baseline for the significant interactions that include these terms. The following figures present the data that is predicted by the model for each significant smooth term of interest.

Figure 3.12 shows the main effect of the %NASDUR on proportions of fixations to the coarticulated image. As expected, participants fixated on the CVN image more when it was partially nasalized (i.e. between 0 and 50N %NASDUR). This is expected as the heard vowel corresponds to the participants' expectations for the realization of a CVN vowel, perhaps less for the 50N stimuli. There is also a decrease in proportions of fixations to the CVN image in the 80N and 100N %NASDUR. This pattern of fixations to the CVN image correspond to a more categorical pattern, since within-category variability on the %NASDUR dimension between 0% and 50% nasalization duration had little influence on the participants' fixation patterns.

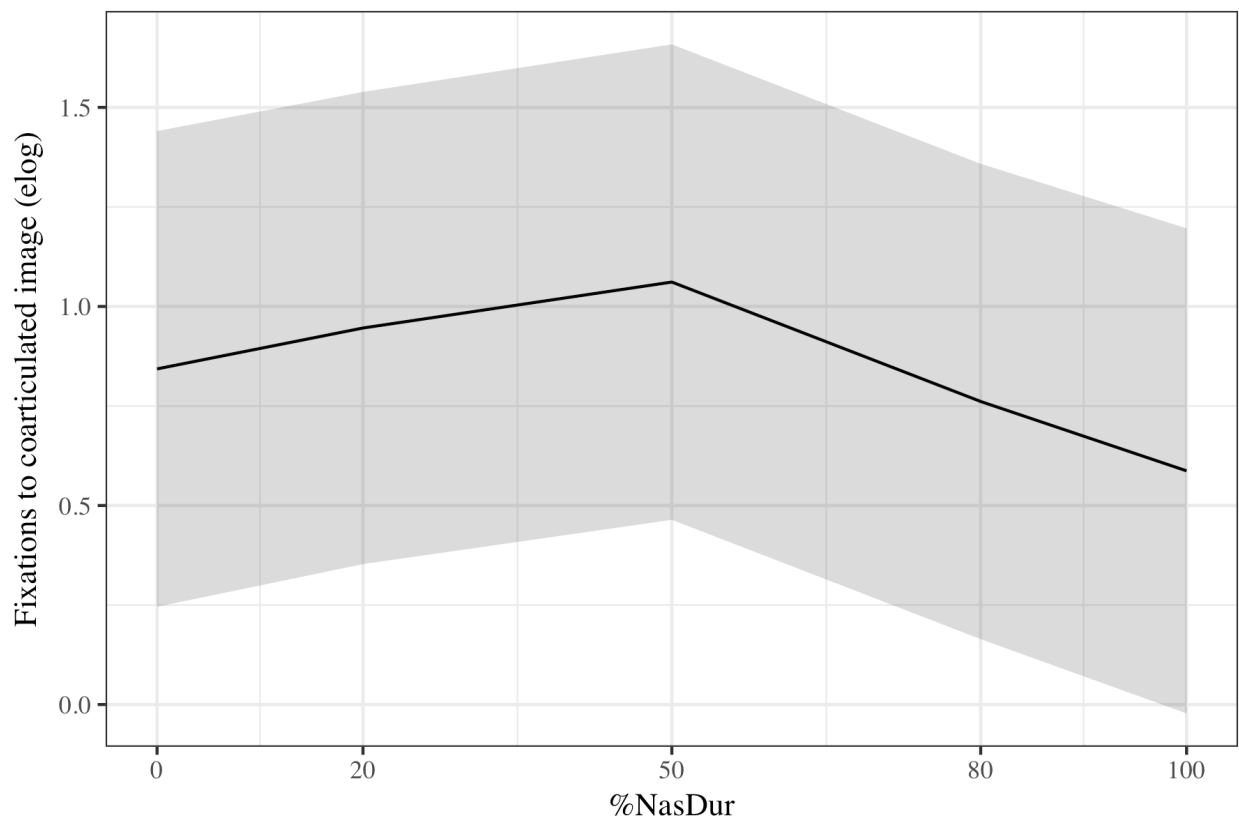


Figure 3.12. Predicted data by %NASDUR – Exp. 1a elog %fix to CVN, CVN choices

In Figure 3.13, the significant interaction between %NASDUR and VOWEL QUALITY through TIME suggests that participants' fixation patterns were variable, based on the quality of the vowel and %NASDUR. For example, fixation patterns for the [o] vowel seem slightly different than the patterns for the other two vowels, especially when the stimuli are nasalized for 50% or more of their duration. This suggests that there may be processing differences based on the quality of the vowel, which can be in part due to the differences in oral articulation underline in Section 1.2.1. Again, the observed patterns of fixations across VOWEL QUALITIES and %NASDUR suggest some complexities in the processing of nasal and coarticulated vowels.

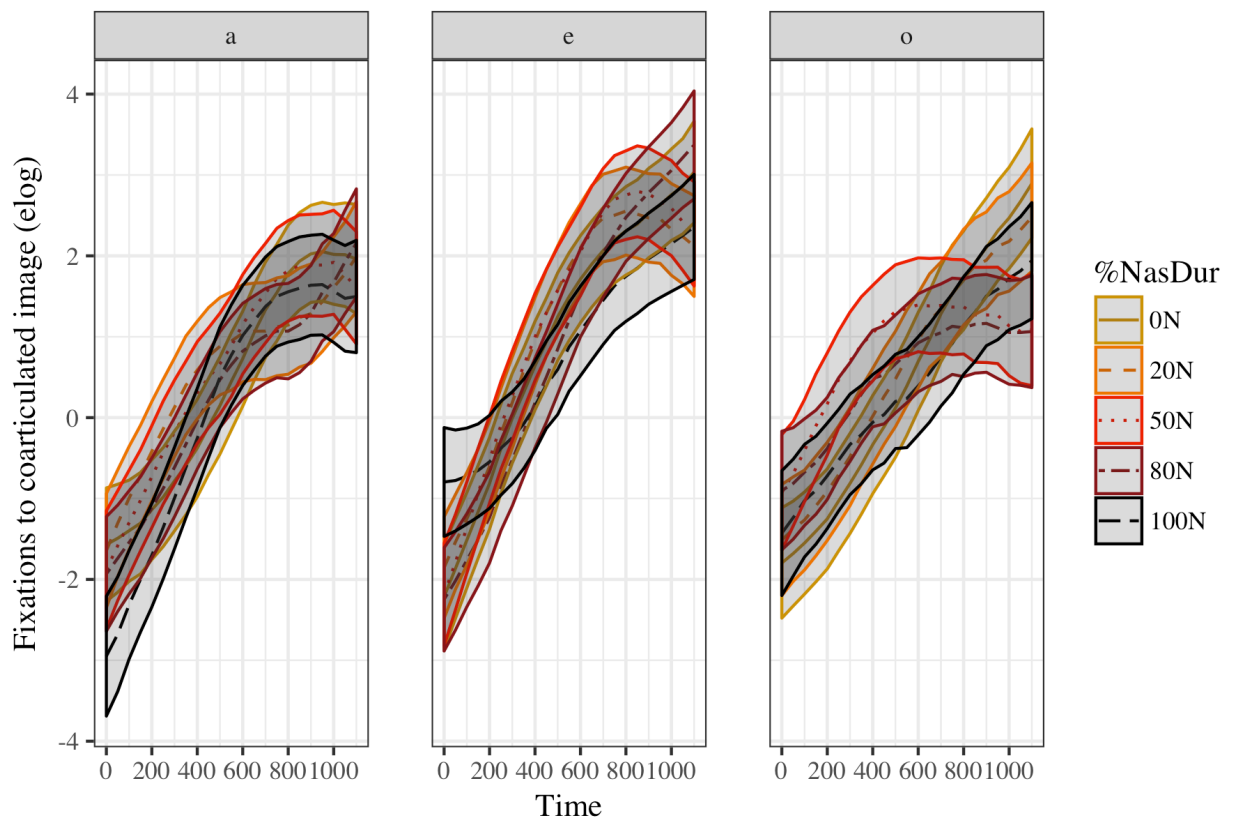


Figure 3.13. Predicted data by %NASDUR and VOWEL QUALITY through TIME – Exp. 1a elog %fix to CVN, CVN choices

3.1.3 *Additional results and discussion*

The results of Experiment 1a indicate that variations in the duration of nasalization have a significant influence on the recognition of phonological nasal ($C\tilde{V}$) and coarticulated nasalized (CVN) vowels by native speakers (L1) of Canadian French (CF). Participants chose the coarticulated (CVN) or nasal ($C\tilde{V}$) image at different rates depending on the condition, and the increase in probability of choosing the nasal ($C\tilde{V}$) item gradually increased as the duration of nasalization on the vowel increased. The analysis of reaction times (RTs) also showed that when a vowel is recognized as a coarticulated nasalized (CVN) vowel, participants are faster in the 0N and 20N conditions than in the other conditions (50N, 80N and 100N), and inversely for vowels that are recognized as phonological nasal ($C\tilde{V}$). Again, this shows differences across conditions, and thus an influence of duration of nasalization on recognition and processing.

Furthermore, as shown by the analysis of proportions of fixations to both the nasal ($C\tilde{V}$) and the coarticulated (CVN) image, the less a vowel is nasalized, the most it is recognized as a coarticulated nasalized vowel through time. Inversely, if a vowel is nasalized for a longer period, it is increasingly recognized as a nasal ($C\tilde{V}$) vowel, even though there is a bias towards the CVN interpretation in the current data. The statistical analyses of chosen images, RTs and proportions of fixations revealed a significant influence of %NASDUR (overall, through TIME, and in interaction with other predictors), thus an influence of nasalization duration on recognition and processing of coarticulated and phonological nasal vowels of CF, which can be gradient (i.e. image choice analysis, and GAMMs on fixations to $C\tilde{V}$).

It also seems like the results of Experiment 1a, when participants chose the nasal ($C\tilde{V}$) image, conform to the gradient recognition hypothesis (McMurray et al., 2002). As observed in Figure 3.9 (proportions of fixations), the influence of duration of nasalization on eye movements is gradient, and not categorical. Results show that shorter proportional duration of nasalization

yields to lower proportions of fixations to the nasal (C $\tilde{V}$) image (i.e. *dent* or *pain*, as opposed to *dame* or *peigne*), while longer duration of nasalization yields to lower proportions of fixations to the nasal (C $\tilde{V}$) image. Importantly, the increase in proportions of fixations to the nasal (C $\tilde{V}$) image is positively correlated to the duration of nasalization values in a gradient manner, which supposes that the acoustic information is gradiently integrated.

On the other hand, when participants chose the coarticulated (CVN) image, patterns of fixations are more categorical. That is, variability in the duration of nasalization between 0N and 50N did not impact fixations to the CVN image, suggesting little sensitivity (or use) to this fine-grained information. Considering the realization of CVN words in Canadian French (Desmeules-Trudel, 2015a; Desmeules-Trudel & Brunelle, 2018), that always have a nasal consonant by definition and for which the vowel can be partially nasalized, it is expected that the 0-to-50N %NASDUR conditions will yield more fixations to the CVN item, if the final nasal appendix is interpreted as an actual nasal consonant.

Additional effects were observed in the statistical analysis of the collected data (see Appendix 3), which the most relevant findings are mentioned here briefly. For example, it is shown that the self-reported PROFICIENCY IN ENGLISH AS A SECOND LANGUAGE (L2) has a continuous effect on RTs (Figure A.2): higher self-reported proficiency leads to shorter RTs in general. This is expected based on the literature on cognitive advantages in more proficient bilinguals (Bialystok, 2015), who can be thought to be more efficient in terms of general information processing, when compared to less proficient bilinguals or monolinguals.¹⁰

¹⁰ Note that the bilingual advantage has been a topic of debate for decades within the literature on L2 acquisition and processing (for a review, see Sabourin and Vīnerte, 2018). However, the current explanation is reasonable regarding the current results, and is raised here as a hypothesis without making any strong claims about the architecture of the bilingual language processing system.

The effect of TRIAL NUMBER (Figure A.1 in Appendix 3) over the course of the experiment was also shown to be significant for both the RT and proportion of fixations analyses. Overall, participants displayed shorter RTs when advancing through the course of the experiment, and also tended to fixate the coarticulated (CVN) image more. This suggests that listeners anticipated CVN stimuli more towards the end of the experiment, although this does not seem entirely linear from the figures. Similarly for proportions of fixations to the CVN image (CVN choices), participants tended to fixate to the CVN image more towards the end of the experiment (both main effect and interaction through TIME; Figure A.7 and Figure A.9).

The final factor that turned out as significant in our analyses is the FREQUENCY OF THE NASAL (C $\tilde{V}$) WORD on the proportions of fixations to the nasal (C $\tilde{V}$) and coarticulated (CVN) image. In the results (Figure A.4 and Figure A.10 in Appendix 3), word pairs in which the nasal (C $\tilde{V}$) word has a higher frequency than other nasal (C $\tilde{V}$) words yield lower proportions of fixations to the nasal (C $\tilde{V}$) image and higher proportions of fixations to the coarticulated (CVN) image, which may seem counterintuitive. In general, higher word frequency leads to higher proportions of fixations to the image corresponding to that word (Dahan et al., 2001a), leading to the expected result that listeners would fixate more the nasal (C $\tilde{V}$) word in our data (as opposed to what is actually found here). However, these results can also be interpreted in light of the inhibition processes during spoken word recognition: if the nasal (C $\tilde{V}$) member of the pair is recognized faster due to its higher frequency value, it can also be inhibited more quickly, therefore allowing the participant to recognize the coarticulated (CVN) word more quickly. This interpretation is speculative, but it is a reasonable way to reconcile the obtained results here with the literature on the topic of the influence of word frequency and word recognition.

Note that the stimuli that were used in Experiment 1a were inherently ambiguous as some of them (50N, 80N and 100N) possessed contradicting cues: a long-nasalized vowel and a short

nasal consonant, a construction that is phonotactically banned according to French phonology. This could have influenced the obtained results and led to effects that would not necessarily be found in circumstances where stimuli were unambiguous. Natural speech in everyday life contains ambiguous word tokens, as coarticulation between surrounding segments is a well-documented phenomenon and because speech conditions are not always ideal, but isolated words do not vary as much (Farnetani & Recasens, 2010). A second experiment was conducted with L1 speakers of CF using unambiguous stimuli to verify if the effects that were found in Experiment 1a also apply when stimuli unambiguously contain a phonological nasal ($C\tilde{V}$) vowel that varies in nasalization duration and is not followed by a nasal consonant/appendix.

3.2 EXPERIMENT 1B

In Experiment 1b, stimuli did not contain conflicting cues, i.e. partly or fully nasalized vowels were not followed by a nasal consonant in the coda. While I found that most phonological nasal vowels have a following nasal appendix (Desmeules-Trudel, 2015a), all vowels were produced in connected speech in this study (within multisyllabic words or followed by another word). In the current work, stimuli are presented in isolation, and therefore nasal ($C\tilde{V}$) and coarticulated (CVN) vowels are expected to be better contrasted. Also, as the stimuli were presented in isolation in the task, it is possible that listeners expected phonological nasal ($\tilde{V}$) vowels not to be followed by a nasal appendix. Furthermore, as mentioned above, informal debriefing with the participants of Experiment 1a suggested that they found the stimuli ambiguous. Stimuli from Experiment 1a were thus modified and disambiguated in Experiment 1b by removing the final 50-ms nasal consonant. This permitted me to test whether words that do

not have conflicting phonetic cues (i.e. variations in nasalization duration only) are processed differently than ambiguous ones.

3.2.1 Methods

3.2.1.1 Participants

Twenty-four native speakers of CF (18 female, 6 male) were paid or received partial course credit for their participation. Twelve speakers were from Québec and twelve from Ontario. They were aged between 17 and 32 years old ($\bar{x} = 21.5$, $SD = 4.4$). All listeners reported normal hearing, normal or corrected-to-normal vision, and did not report any type of language, hearing or speech impairment. All participants completed a background language questionnaire (see Appendix 1), and self-reported knowing English as a second language at a moderate to high level of proficiency (see Figure 3.14), as well as other languages. Five were early bilinguals, and nineteen learned English as a second language at a later age. Importantly, no speaker had been exposed to another language containing contrastive nasal vowels in its phonological inventory.

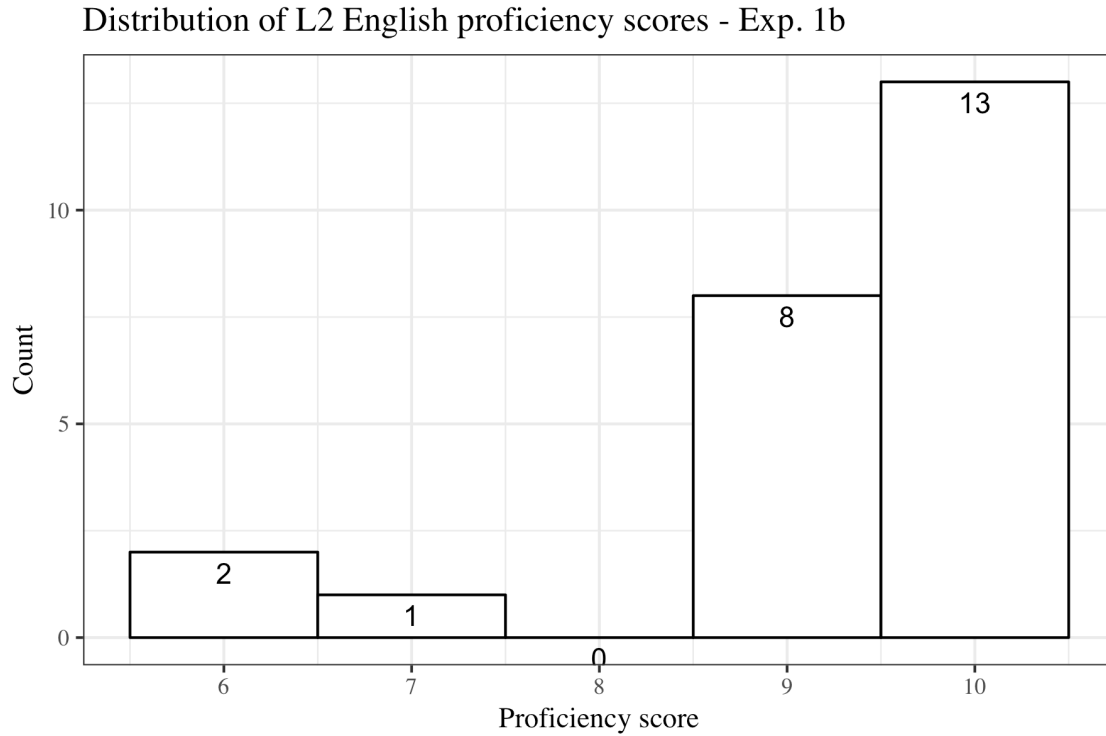


Figure 3.14. Distribution of self-reported L2 English proficiency scores out of 10 for Exp. 1b participants

3.2.1.2 Stimuli

Stimuli were the same word triads as in Experiment 1a (Table 3.1). They were paired with the same images used in the previous experiment, and the same recordings and cross-spliced stimuli were used. The only modification that was performed on the pre-spliced stimuli was the removal of the final 50-ms nasal appendix (Figure 3.15). The same duration of nasalization conditions (0N, 20N, 50N, 80N and 100N) were used, and are summarized in Table 3.10.

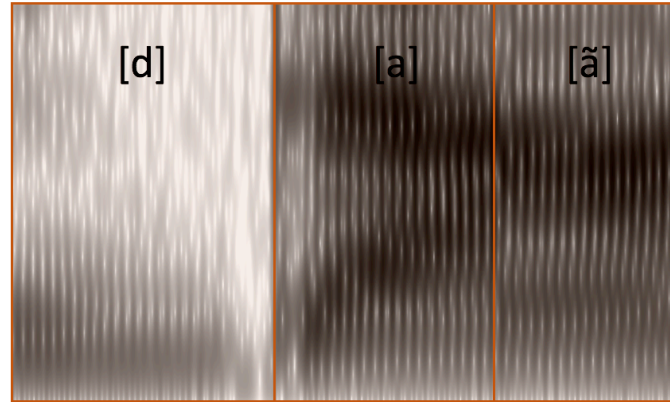


Figure 3.15. Spectrogram of *dame/dent* [dam]/[dã] ‘lady/tooth’ stimulus in the 50N %NASDUR with no nasal appendix – Exp. 1b

Table 3.10

Summary of the experimental conditions (%NASDUR) – Exp. 1b

→ Vowel duration →			
Condition	0N	Oral (100%)	
	20N	Oral (80%)	Nasal (20%)
	50N	Oral (50%)	Nasal (50%)
	80N	Oral (20%)	Nasal (80%)
	100N	Nasal (100%)	

3.2.1.3 Procedure and analysis

The experimental procedure, including calibration, familiarization, stimuli presentation, and collected data was the same as Experiment 1a. The analysis (GAMMs) and investigated variables were the same as Experiment 1a, and tables summarizing the statistical predictors are presented before the results.

3.2.2 Results

This section presents the results of Experiment 1b, focusing on the chosen images and reaction times (RTs) in Section 3.2.2.1. The eye tracking results are presented in Section 3.2.2.2. Recall that the chosen image measure below corresponds to the image that was fixated the longest in the last 1000 ms of each trial.

3.2.2.1 Chosen images and reaction times

In Experiment 1b (Figure 3.16), listeners chose the nasal (C $\tilde{V}$) image in a greater proportion (between 36.48% and 94.26%) than the coarticulated (CVN) image (between 3.83% and 19.42%; see upper panel). Also note that in the 0N condition, participants chose the oral (CVC) image in 44.09% of cases. This is due to the cross-splicing procedure (see Section 3.1.1.2). In the 0N condition, the vowel from a CVC item was pasted within the CVN matrix, but the formant transitions at the end of the vowel led to a following oral consonant with a place of articulation that was sometimes different from the nasal consonant in the CVN word (e.g. *dame* [dam] ‘lady’ ~ *datte* [dat] ‘date (fruit)’). As the nasal consonant from the CVN word was fully removed in Experiment 1b, participants who were presented with a fully oral vowel containing transitions towards the oral consonant of the CVC item chose this latter word as the best fit in this condition. The remainder of the presentation and GAMM analysis (bottom panel in Figure 3.16 and Table 3.11) will focus on the nasal (C $\tilde{V}$) and coarticulated (CVN) items. For the chosen image analysis, only the C $\tilde{V}$ and CVN chosen images were kept (1826/2160 trials), getting rid of the CVC and filler (i.e. incorrect) responses.

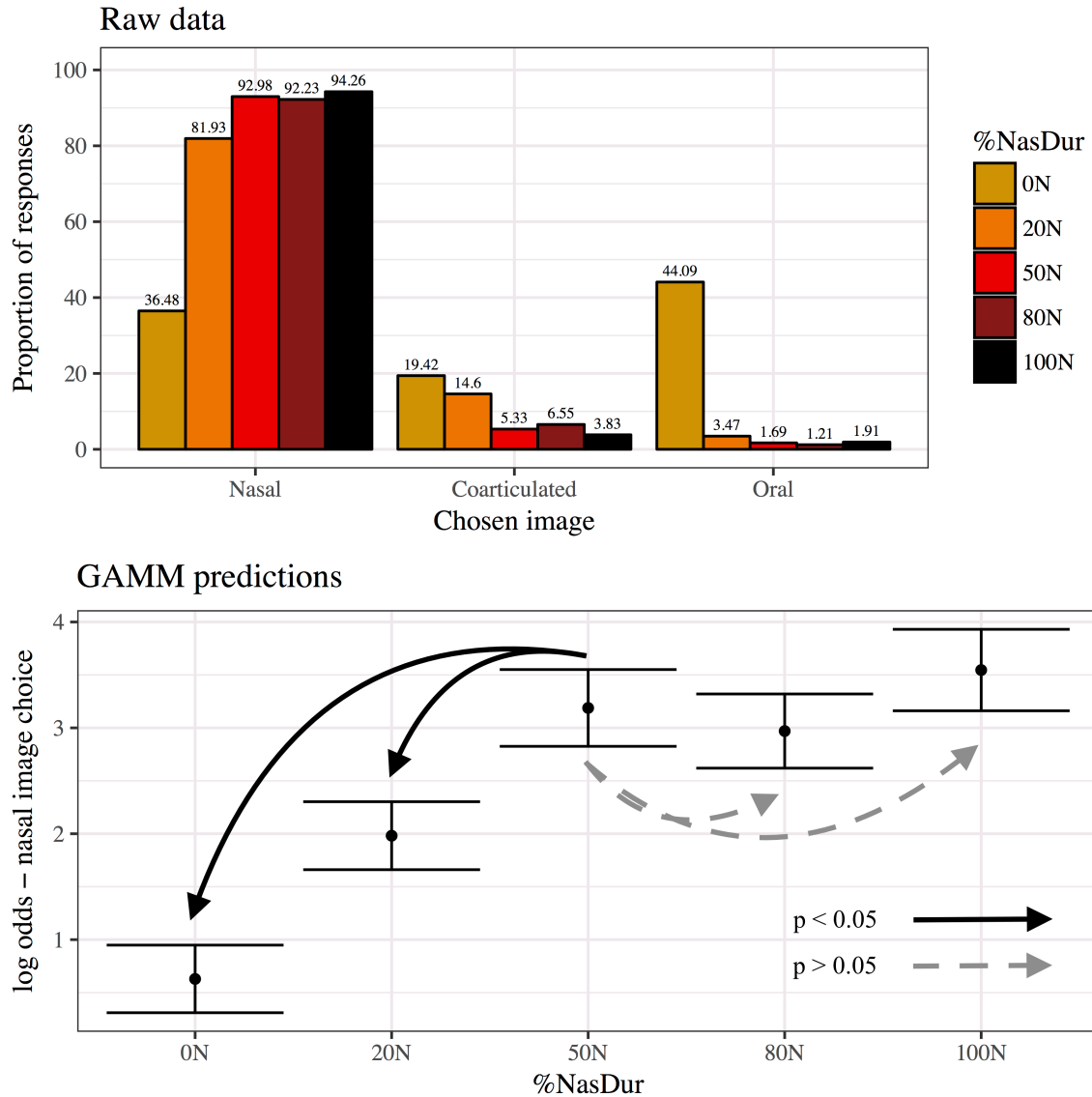


Figure 3.16. Proportions of image choices for nasal (C $\tilde{V}$) and coarticulated (CVN) images (upper panel) and GAMM predictions for image choice analysis (i.e. probability of C $\tilde{V}$ image choice by %NASDUR) – Exp. 1b

The GAMM analysis suggests that when compared to the 50N baseline, the probability of choosing the C $\tilde{V}$ image in the 50N %NASDUR condition is significantly different (see Table 3.4) from the 0N and 20N levels of %NASDUR, but not from 80N and 100N. This suggests that the probability of choosing the C $\tilde{V}$ image gradually increased between 0% and 50% %NASDUR, and remained stable across 50%, 80% and 100% nasalization duration. This suggest a categorical

pattern of perception, were within-category variability between 50% and 100% nasalization duration did not impact recognition, which can be interpreted as a ceiling effect. Additional effects are illustrated in Appendix 3.

Table 3.11

GAMM results – Exp. 1b image choices

A. parametric coefficients	Estimate	Std. Error	t-value	p-value	
Intercept	2.6775	0.3333	8.0344	< 0.0001	***
Vowel[e]	1.7958	0.3330	5.3933	< 0.0001	***
Vowel[o]	0.1697	0.3771	0.4501	0.6526	
L2_earlyExp	-0.6155	0.3108	-1.9806	0.0476	*
%NasDur[0N]	-2.9229	0.2909	-10.0467	< 0.0001	***
%NasDur[20N]	-1.2174	0.2695	-4.5172	< 0.0001	***
%NasDur[80N]	-0.2008	0.3032	-0.6623	0.5077	
%NasDur[100N]	0.3745	0.3424	1.0937	0.2741	
B. smooth terms	edf	Ref.df	F-value	p-value	
s(Freq_coart)	4.9107	5.4870	69.5341	< 0.0001	***
Random_Participant	13.7431	22.0000	35.4337	0.0001	***
Random_Item	0.0022	5.0000	0.0016	0.2073	

RTs were a lot shorter when participants chose the nasal (C $\tilde{V}$) image than when they chose the coarticulated (CVN) image overall, as seen in Figure 3.17. When participants choose the nasal (C $\tilde{V}$) image (left), their RTs got even smaller as the duration of nasalization (i.e. %NASDUR) increased. No clear pattern seemed to emerge when participants chose the coarticulated (CVN) image. Note that the GAMM analysis below did not find a significant effect of CHOSEN IMAGE (main effect or within an interaction), suggesting that the number of CVN tokens (i.e. 198/1826) was too small to have an effect on RT values overall.

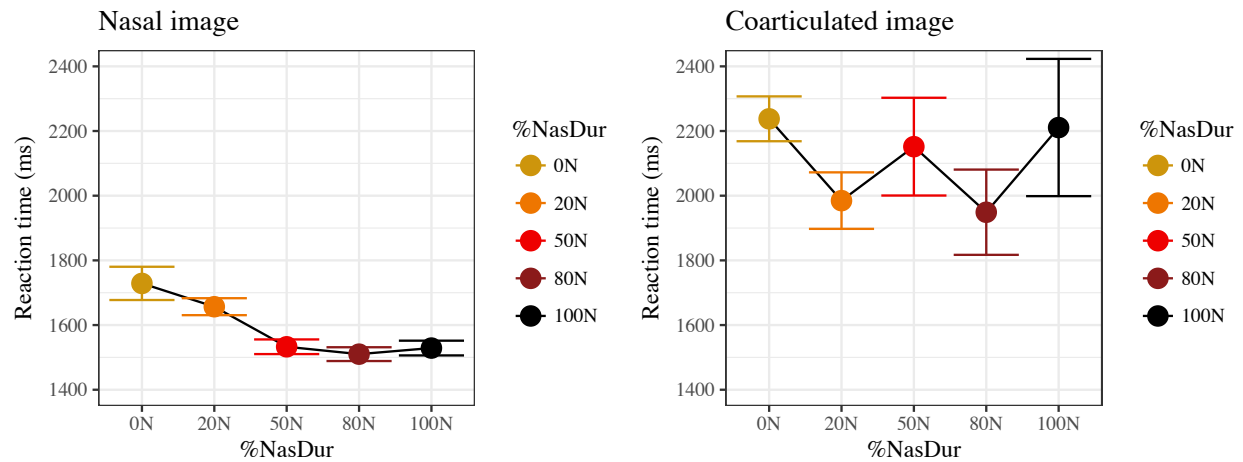


Figure 3.17. RT averages for nasal (C $\tilde{V}$) image choices by %NASDUR (left) and coarticulated (CVN) image choices (right) by %NASDUR – Exp. 1

The remainder of this subsection presents the statistical analysis (GAMMs) conducted on log-transformed RTs. Maximum-likelihood chi-square tests on the fully specified model yielded significant random smooths by PARTICIPANTS and %NASDUR, and by ITEM and %NASDUR. Predictors of interest and interactions that were assessed are summarized in Table 3.12.

Table 3.12

Summary of the independent variables for the full RT GAMM – Exp. 1b

Parametric terms	Levels
CHOSEN IMAGE	Factor w/ 2 levels: Nasal, Coarticulated image
VOWEL QUALITY	Factor w/ 3 levels: [a], [e], [o]
PROVINCE OF ORIGIN	Factor w/ 3 levels: NB, ON, QC
EXPOSURE TO L2 BEFORE 5	Factor w/ 2 levels: True or False
Smooth terms	Levels
%NASDUR	Numerical (centered): -50 to 50
TRIAL NUMBER	Numerical: 1 to 160
LOG FREQUENCY (CVN)	Numerical: -1.0804 to 1.1756
LOG FREQUENCY (C $\tilde{V}$)	Numerical: -0.5229 to 0.7491
ENGLISH PROFICIENCY	Numerical: 7 to 10
Interactions	
%NASDUR*VOWEL QUALITY	
%NASDUR*FIXATED IA	
Random effects	
Smooth term:	
PARTICIPANTS*%NASDUR	
Smooth term:	
MINIMAL PAIR LABEL*%NASDUR	

Results of the final model are presented in Table 3.13 with CHOSEN IMAGE and VOWEL QUALITY as parametric factors (the nonsignificant VOWEL QUALITY factor was kept in as a reference for the smooth parameters), significant smooth terms for %NASDUR and the 2-way interaction between %NASDUR and VOWEL QUALITY (although see below for more details on the interaction significance). Presentation of the results will focus on the significant %NASDUR factor and %NASDURXVOWEL QUALITY interaction, as they are crucial to the research question.

Parametric coefficients in Table 3.13 suggest that overall, when participants fixated on the coarticulated (CVN) image, RTs are significantly different from when participants fixated the nasal image (3.254 ± 0.065 , $t = -7.332$ $p < 0.05$). This seems to match the raw RT values presented in Figure 3.17. The following figures (3.14 and 3.15) present the data that is predicted by the model according to each significant smooth term of interest.

Table 3.13

GAMM results – Exp. 1b RTs

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
Intercept	3.2542	0.0298	109.1077	< 0.0001
Chosen_image	-0.0654	0.0089	-7.3322	< 0.0001
Vowel_[e]	-0.0067	0.0380	-0.1762	0.8601
Vowel_[o]	0.0070	0.0381	0.1823	0.8554
B. smooth terms	edf	Ref.df	F-value	p-value
s(%NasDur)	2.2968	2.5042	5.8814	0.0009
s(%NasDur):Vowel_[a]	1.7633	1.9777	2.4848	0.0672
s(%NasDur):Vowel_[e]	1.0004	1.0006	2.3193	0.1280
s(%NasDur):Vowel_[o]	0.4313	0.5880	0.1285	0.7835
Random_Participant	27.2236	94.0000	3.9408	< 0.0001
Random_Item	16.4472	30.0000	7.5567	< 0.0001

Figure 3.18 shows the RT data predicted by the final GAMM, suggesting a significant decrease in RTs with the increase in %NASDUR overall. Considering that the nasal ($C\tilde{V}$) image was chosen the most in Experiment 1b (i.e. 1628/1826 times, for $C\tilde{V}$ and CVN responses only), this suggests that an increase in nasalization duration causes lower processing load for listeners. The absence of a significant interaction on RTs between the CHOSEN IMAGE and %NASDUR is probably due to the low number of CVN choices. Therefore, the decrease in RTs is consistent with the realization of vowel nasalization in CF; phonological nasal vowels ($\tilde{V}$) are nasalized for a longer period than coarticulated (VN) vowels, therefore interpreting stimuli that were nasalized for a long period as nasal $\tilde{V}$ s requires less processing than interpreting them as coarticulated VNs.

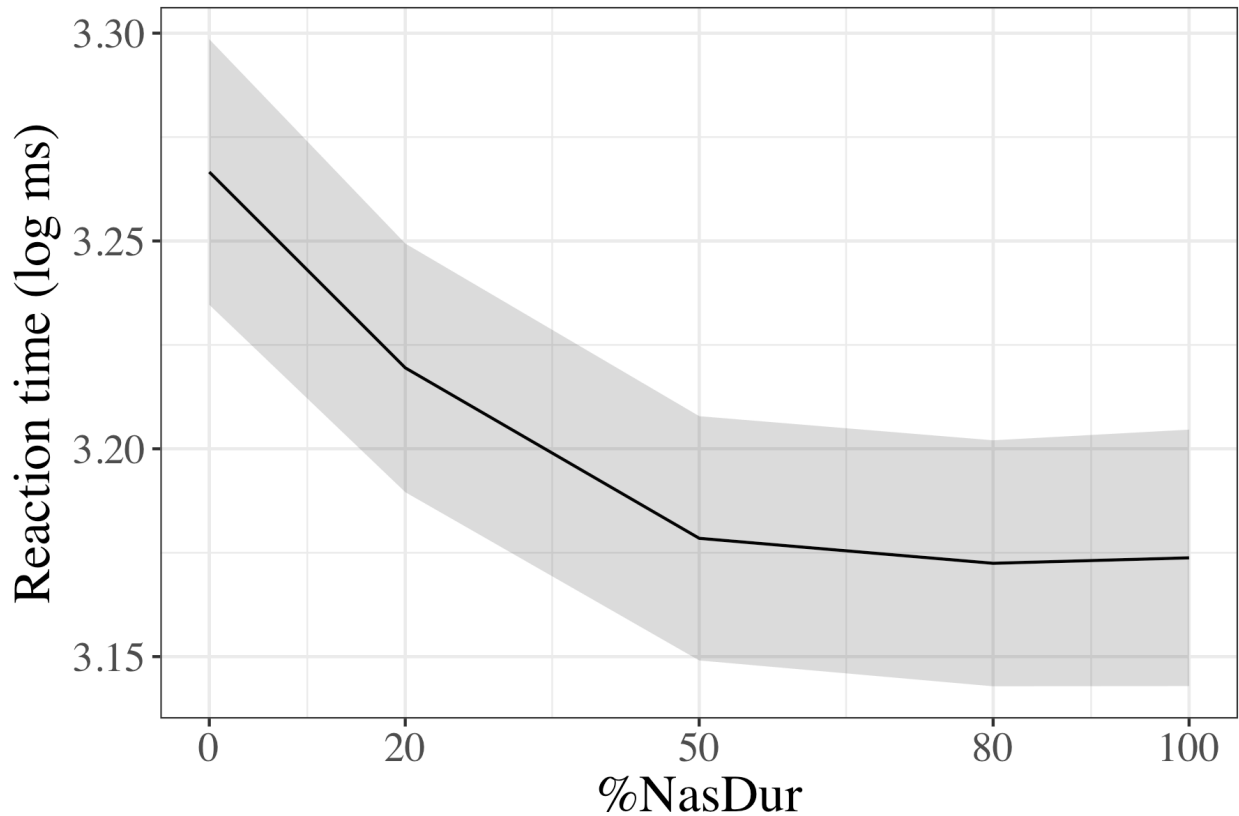


Figure 3.18. Predicted data by %NASDUR – Exp. 1b RTs

The predicted data presented in Figure 3.19 below show the significant interaction (note that $p = 0.067$ for the [a] level of the interaction, suggesting a trend towards significance) of CONDITION and VOWEL QUALITY. It is shown that the [o] vowel in the 0N and 20N conditions yielded longer RTs than the other two vowel qualities. RTs also seem to decrease more for [o] than the other vowels, as the latter seem to slightly increase between 50N, 80N and 100N for [a] and [e]. This suggests differences across vowel qualities in terms of processing, and these will be discussed below in 3.2.3.

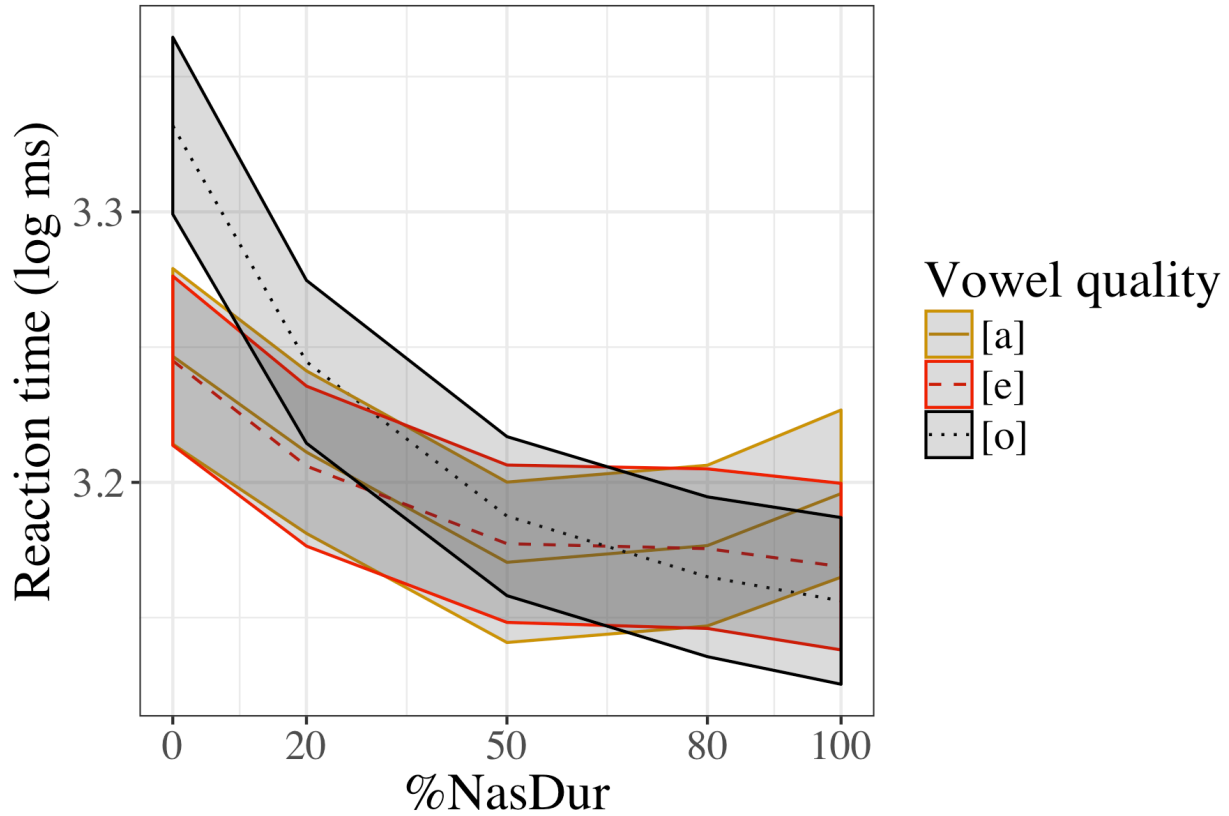


Figure 3.19. Predicted data by %NASDUR and VOWEL QUALITY – Exp. 1b RTs

3.2.2.2 Eye movements

In Experiment 1b, participants chose the C $\tilde{V}$ image 1628 times overall, and chose the CVN image 198 times overall. The time window extended from 200 ms after vowel onset to account for the eye movement programming delay (Fischer, 1992), up to 1100 ms after vowel onset based on the peak in fixations to the target in the 100N condition. Figure 3.20 presents patterns of fixations to the C $\tilde{V}$ (leftmost panels) and CVN images (rightmost panels) when listeners chose the C $\tilde{V}$ image (upper panels) and when they chose the CVN image (lower panels).

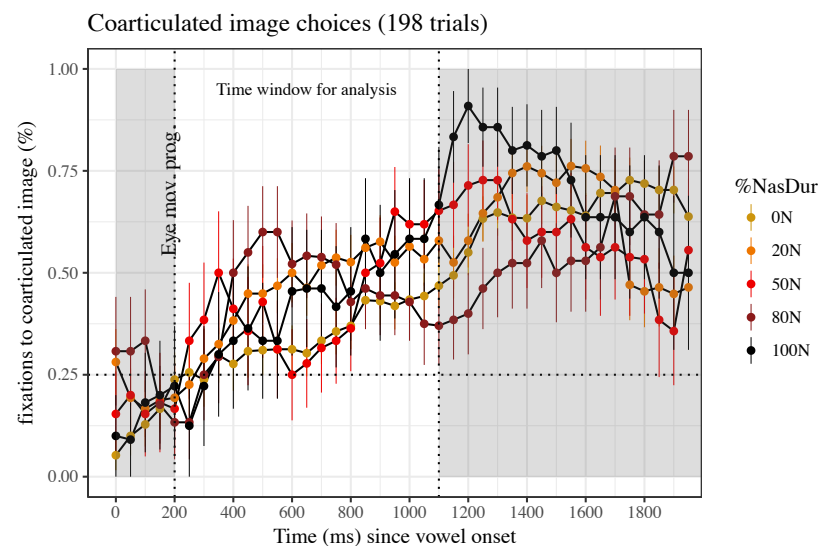
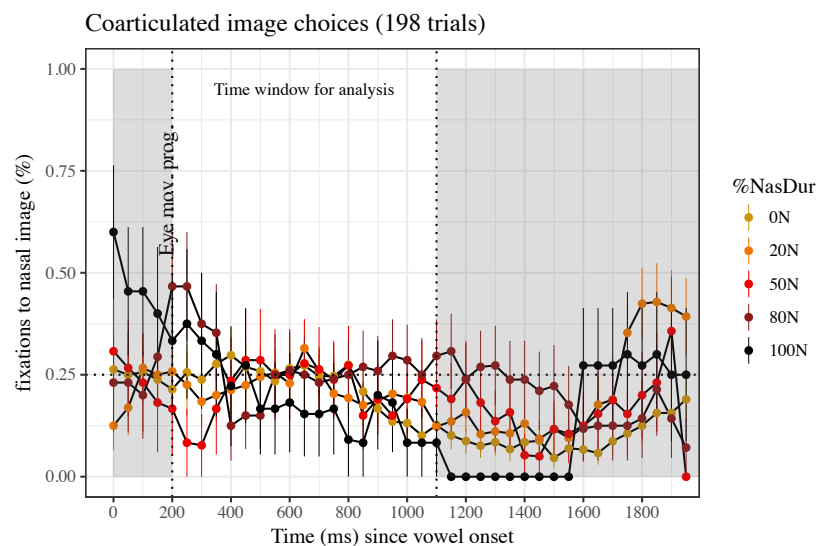
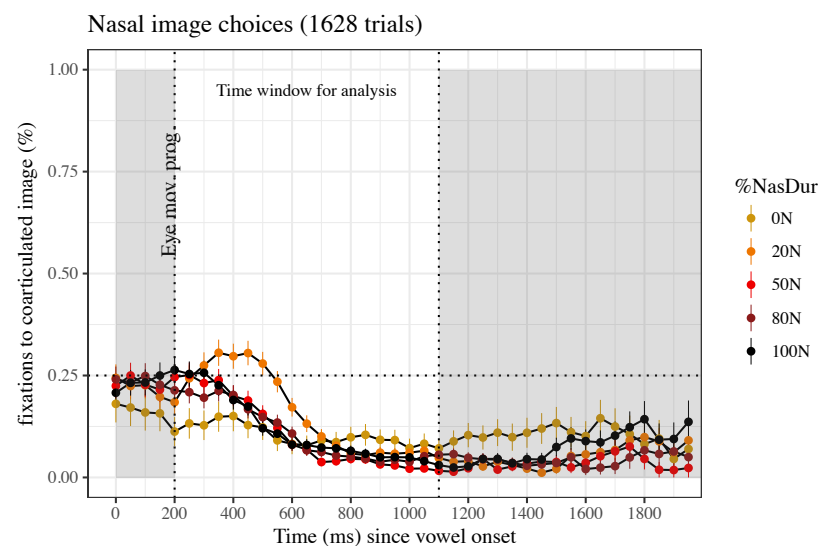
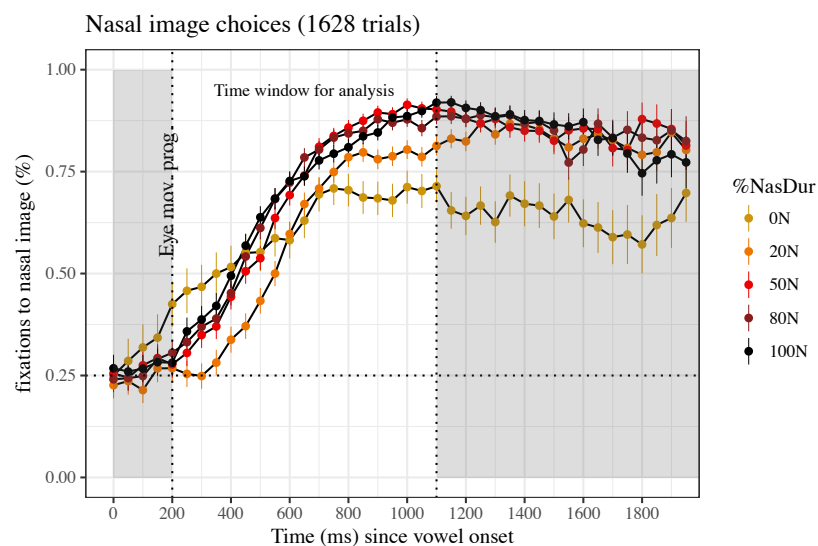


Figure 3.20. Proportions of fixations to the nasal ($C\tilde{V}$) image (leftmost) and coarticulated (CVN) image (rightmost), nasal ($C\tilde{V}$) choices (upper) and coarticulated (CVN) choices (lower) – Exp. 1b

Looking at fixation patterns in Figure 3.20. Proportions of fixations to the nasal ($C\tilde{V}$) image (leftmost) and coarticulated (CVN) image (rightmost), nasal ($C\tilde{V}$) choices (upper) and coarticulated (CVN) choices (lower) – Exp. 1b (upper left), it seems like there is a positive relationship between the proportions of fixations to the $C\tilde{V}$ image and the duration of nasalization on the vowel throughout the time window when listeners chose the $C\tilde{V}$ image, although not as markedly as in Experiment 1a. This pattern seems consistent with the image choice analysis in Section 3.2.2.1. The models were fitted to the eye-tracking data in a similar way as in Experiment 1a. The predictors are summarized in Table 3.14.

Table 3.14

Summary of the independent variables for the full GAMM – Exp. 1b %fix

Parametric terms	Levels
CHOSEN IMAGE	Factor w/ 2 levels: Nasal (baseline), Coarticulated image
VOWEL QUALITY	Factor w/ 3 levels: [a], [e], [o]
PROVINCE OF ORIGIN	Factor w/ 2 levels: ON, QC
EXPOSURE TO L2 BEFORE 5	Factor w/ 2 levels: True or False
Smooth terms	Levels
TIME	Numerical: 0 to 900, separated in 50 ms time bins
%NASDUR	Numerical (centered): -50 to 50
TRIAL NUMBER	Numerical: 1 to 160
LOG FREQUENCY (CVN)	Numerical: -1.0804 to 1.1756
LOG FREQUENCY (C $\tilde{V}$)	Numerical: -0.5229 to 0.7491
ENGLISH PROFICIENCY	Numerical: 6 to 10
Interactions	
TIME* FIXATED IA	
TIME* VOWEL QUALITY	
TIME* PROVINCE OF ORIGIN	
TIME* EXPOSURE TO L2 BEFORE 5	
TIME*%NASDUR	
TIME* TRIAL NUMBER	
TIME* LOG FREQUENCY (COART.)	
TIME* LOG FREQUENCY (NASAL)	
TIME* ENGLISH PROFICIENCY	
TIME*%NASDUR *VOWEL QUALITY	
TIME*CONDITION*FIXATED IA	
Random effects	
Smooth term: PARTICIPANTS*%NASDUR	
Smooth term:	
MINIMAL PAIR LABEL *%NASDUR	

Results of the first GAMM (fixations to C $\tilde{V}$ image, C $\tilde{V}$ choices; 1628 trials) are presented in Table 3.15 with VOWEL QUALITY as a parametric factor, significant smooth terms for TIME, and %NASDUR, as well as 2-way interactions between TIME and TRIAL NUMBER, and %NASDUR, and a 3-way interactions between TIME, %NASDUR, and VOWEL QUALITY. Presentation of the results will focus on the significant %NASDUR factor, and the TIMEX%NASDUR and TIMEX%NASDURXVOWEL QUALITY interactions. Illustration of the additional significant smooths

(TIME and TIME_XTRIAL NUMBER) is found in Appendix 3. Auto-correlation AR1 value was empirically determined to 0.815 based on the data.

Table 3.15

GAMM results – Exp. 1b %fix to C $\tilde{V}$, C $\tilde{V}$ choices

A. parametric coefficients	Estimate	Std. Error	t-value	p-value	
Intercept	1.5422	0.3116	4.9495	< 0.0001	***
Vowel_[e]	-0.1334	0.3565	-0.3742	0.7083	
Vowel_[o]	-0.0679	0.3584	-0.1895	0.8497	
B. smooth terms	edf	Ref.df	F-value	p-value	
s(Time)	5.5634	6.0726	20.4748	< 0.0001	***
s(Trial)	1.0055	1.0110	2.0195	0.1548	
s(%NasDur)	2.4202	2.7570	31.7953	< 0.0001	***
s(Time:Trial)	11.2085	13.7047	2.9644	0.0002	***
s(Time:%NasDur)	10.4460	12.7935	4.0487	< 0.0001	***
s(Time:%NasDur):Vowel_[a]	9.7109	12.0075	2.7607	0.0009	***
s(Time:%NasDur):Vowel_[e]	1.0413	1.0619	1.7953	0.1779	
s(Time:%NasDur):Vowel_[o]	8.2731	10.7526	3.1350	0.0003	
Random_Participant	131.9633	215.0000	2.6973	< 0.0001	***
Random_Item	58.3876	78.0000	6.1658	< 0.0001	***

The VOWEL QUALITY factor was kept in as a reference intercept for the smooths containing the VOWEL QUALITY factor in the B section of Table 3.15 (Wood, 2017). The following figures present the data that is predicted by the model for each significant smooth term of interest.

The %NASDUR participants were presented with also has a significant impact on the proportions of fixations to the target, as shown in Figure 3.21. Between 0N and 50N, proportions of fixations to the nasal (C $\tilde{V}$) item increase dramatically, suggesting that an increase in duration of nasalization when the vowel is not nasalized or only for a short period makes a difference in proportions of fixations. However, when the vowel is nasalized for 50% or more of its duration (50N, 80N and 100N conditions), the differences in proportions of fixations to the nasal (C $\tilde{V}$)

image between %NASDUR seem somewhat minimal. This is expected based on the realization of CF vowel nasalization, as phonological nasal ($\tilde{V}$) vowels are nasalized in production for at least 50% of their duration. This pattern could be partially consistent with the gradient recognition hypothesis, as the gradient increase in %NASDUR seems to have an influence on proportions of fixations, until 50% nasalization duration is reached. However, it remains possible that listeners interpreted the 20N condition as a poor instance of a $C\tilde{V}$ word (it was nasalized only for a short period), which could explain the overall lower proportions of fixations. Participants may thus have *categorically* recognized unambiguous stimuli (20N or more %NASDUR).

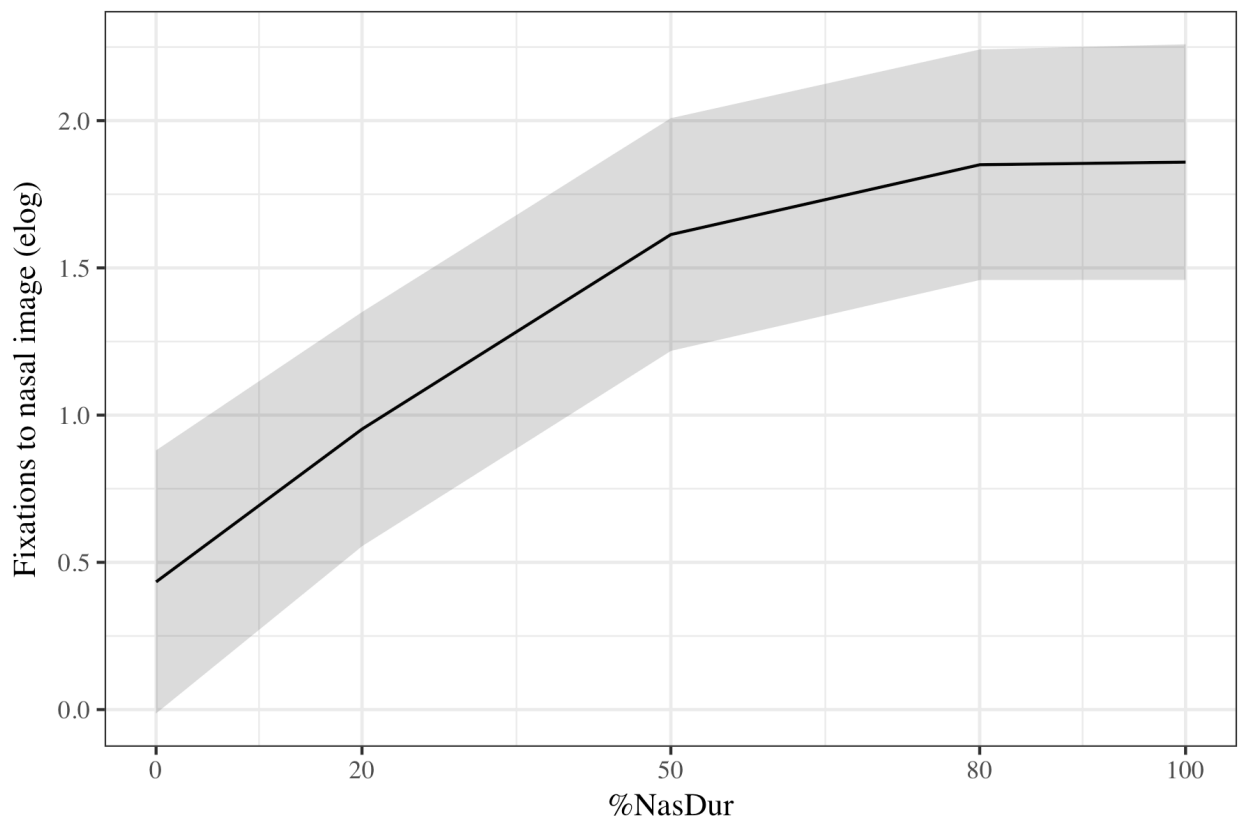


Figure 3.21. Predicted data by %NASDUR – Exp. 1b elog %fix to $C\tilde{V}$, $C\tilde{V}$ choices

Figure 3.22 presents the predicted data, focusing on the significant interaction between TIME and %NASDUR. The predicted data suggests that overall, in the 0N %NasDur, participants fixated less to the nasal (C $\tilde{V}$) image over a significant period of time within the analysis window. Furthermore, the 20N %NASDUR condition also yielded slightly less fixations to the nasal (C $\tilde{V}$) image, and the other three %NASDUR values behaved similarly. This is consistent with the raw data presented in Figure 3.20 (i.e. upper left panel), suggesting a categorical/ceiling effect of %NASDUR on fixations to the nasal (C $\tilde{V}$) image.

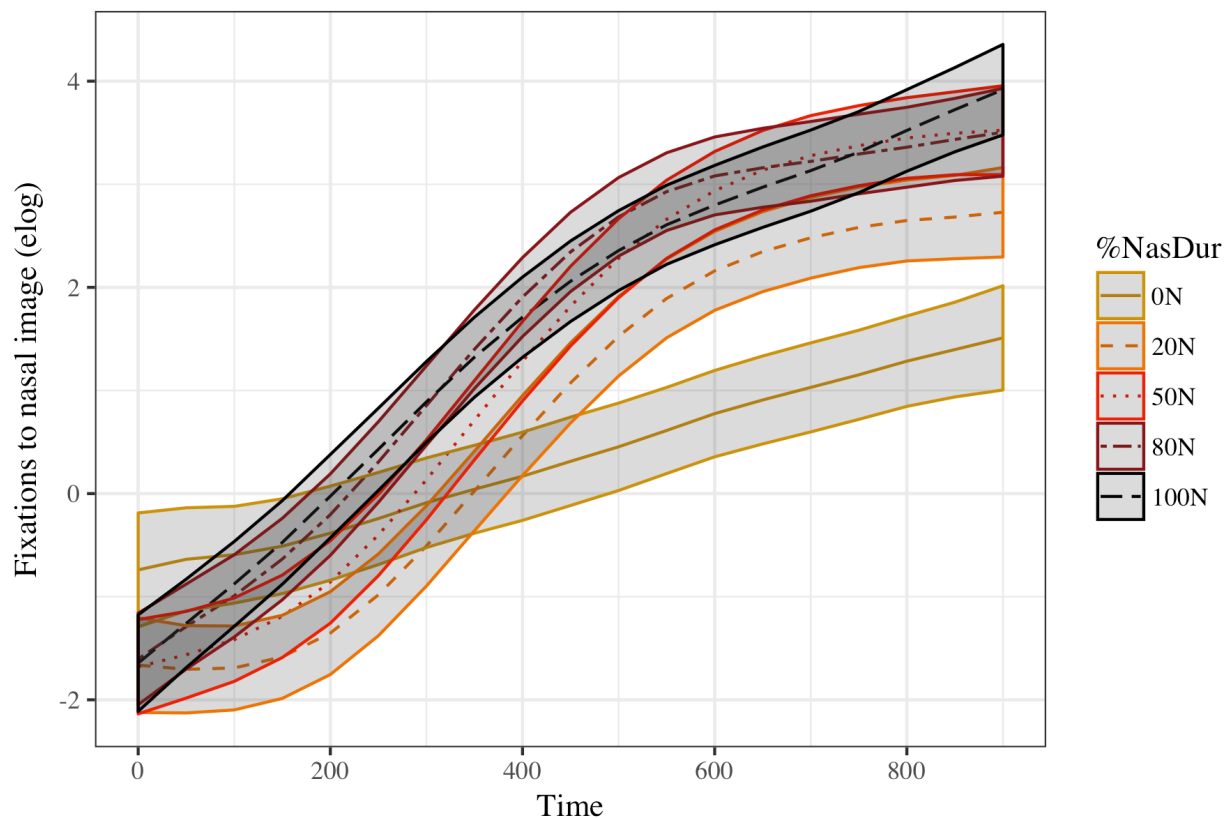


Figure 3.22. Predicted data by %NASDUR through TIME – Exp. 1b e-log %fix to C $\tilde{V}$, C $\tilde{V}$ choices

The three-way interaction between TIME, VOWEL QUALITY and %NASDUR is illustrated in Figure 3.23. For the [a] vowel in the 0N %NASDUR, there seems to be an increase in proportions

of fixations to the nasal (C $\tilde{V}$) image between 0 and 200 ms, a decrease between 200 and 500 ms, and an increase between 500 ms and time window offset. The other four %NASDUR seem to behave relatively similarly across vowel qualities, therefore suggesting that the 0N (and 20N, to some extent, due to the different curve shapes) %NASDUR is mainly at the origin of the statistical significance of the three-way interaction. Also note that the only significant smooth for this interaction concerns the [a] level of the VOWEL QUALITY factor (Table 3.15).

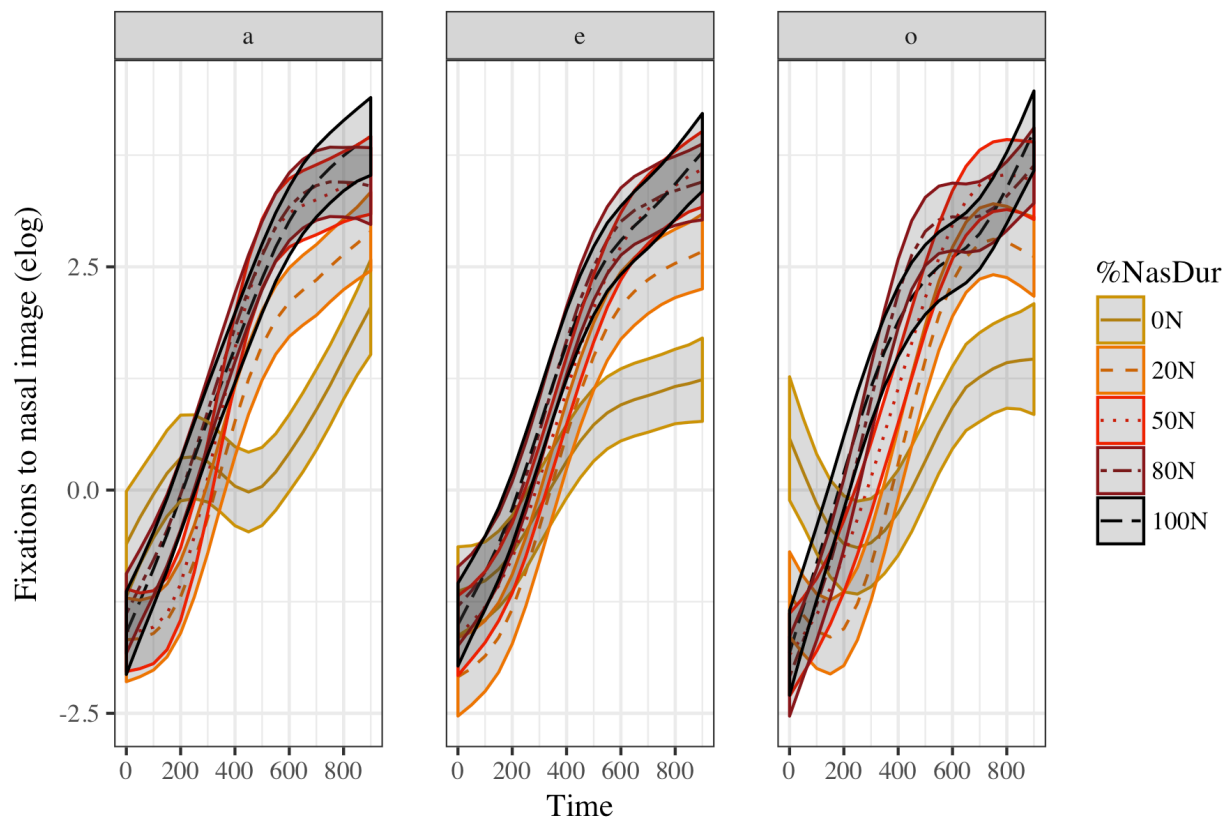


Figure 3.23. Predicted data by %NASDUR and VOWEL QUALITY through TIME – Exp. 1b elog %fix to C $\tilde{V}$, C $\tilde{V}$ choices

In the second GAMM, I modeled the (empirical-logit-transformed) proportions of fixations to the coarticulated (CVN) image when participants chose the CVN image. However,

these analyses are based on only 198 trials, therefore the significant factor must be interpreted with care. Results of the model are presented in Table 3.16. Presentation of the results will focus on the significant TIMEXVOWEL and TIMEXNASDURXVOWEL interactions. Additional effects (TRIAL, FREQUENCY (COARTICULATED), TIMEX TRIAL, TIMEXFREQUENCY (NASAL)) are illustrated in Appendix 3. Auto-correlation AR1 value was determined to 0.809 based on the data.

Table 3.16

GAMM results – Exp. 1b %fix to CVN, CVN choices

A. parametric coefficients	Estimate	Std. Error	t-value	p-value	
Intercept	-1.0498	0.5645	-1.8596	0.0630	
Vowel_[e]	0.2898	0.6048	0.4792	0.6318	
Vowel_[o]	0.5117	0.5830	0.8777	0.3802	
ProvinceQC	0.5308	0.6660	0.7969	0.4255	
B. smooth terms	edf	Ref.df	F-value	p-value	
s(Time)	3.6774	4.3632	1.8873	0.1057	
s(Trial)	6.1412	7.2598	4.0930	0.0001	***
s(%NasDur)	1.1766	1.3188	1.7197	0.2240	
s(log_freqCoart.)	1.6844	1.8999	3.1122	0.0646	
s(log_freqNasal)	2.0930	2.3939	2.4688	0.2519	
s(Time):Vowel_[a]	1.0041	1.0073	0.0212	0.8874	
s(Time):Vowel_[e]	0.0120	0.0209	0.0031	0.9935	
s(Time):Vowel_[o]	5.4947	6.7072	2.7645	0.0084	**
s(Time:Province_ON)	0.0019	0.0024	0.0043	0.9974	
s(Time:Province_QC)	3.9708	4.7167	1.6218	0.1391	
s(Time:Trial)	7.7755	9.8184	3.3560	0.0003	***
s(Time:%NasDur)	2.8350	3.3239	0.6372	0.5334	
s(Time:Freq_nasal)	2.4779	2.9626	4.1467	0.0061	**
s(Time:%NasDur):Vowel_[a]	7.7313	10.0901	2.2087	0.0146	*
s(Time:%NasDur):Vowel_[e]	7.5013	10.0475	1.8359	0.0463	*
s(Time:%NasDur):Vowel_[o]	7.0572	9.4198	1.9691	0.0391	*
Random_Participant	95.4212	214.0000	1.2070	< 0.0001	***
Random_Item	3.4895	76.0000	0.0488	0.1645	

The VOWEL QUALITY and PROVINCE factors were kept in the model to include intercepts for the smooth terms in B. The non-significant TIME, %NASDUR, FREQUENCY (NASAL) factors and

`TIMEX%NASDUR` interaction were also kept in to include a reference for higher-order interactions. Also note that the non-significant `FREQUENCY (COARTICULATED)` and `TIMEXPROVINCE` effects are also included in the model, since they were significant in the initial model fitted with the maximum-likelihood method; the presented model was fitted with the restricted maximum-likelihood method (Porretta et al., 2016).

Figure 3.24 shows the predicted data based on the significant interaction of `VOWEL QUALITY` through `TIME`. The curve shapes suggest some differences among vowel qualities. For example, fixations to the CVN image increase between 300 ms and 700 ms for the [o] items (black dotted line), while they seem more stable for the same time interval for [a] and [e] items. Although it is difficult to find a definite explanation as to why such patterns emerged, it is possible that different vowel qualities are processed differently, regardless of the nasalization status, and further research will be necessary to disentangle the factors of influence in this case.

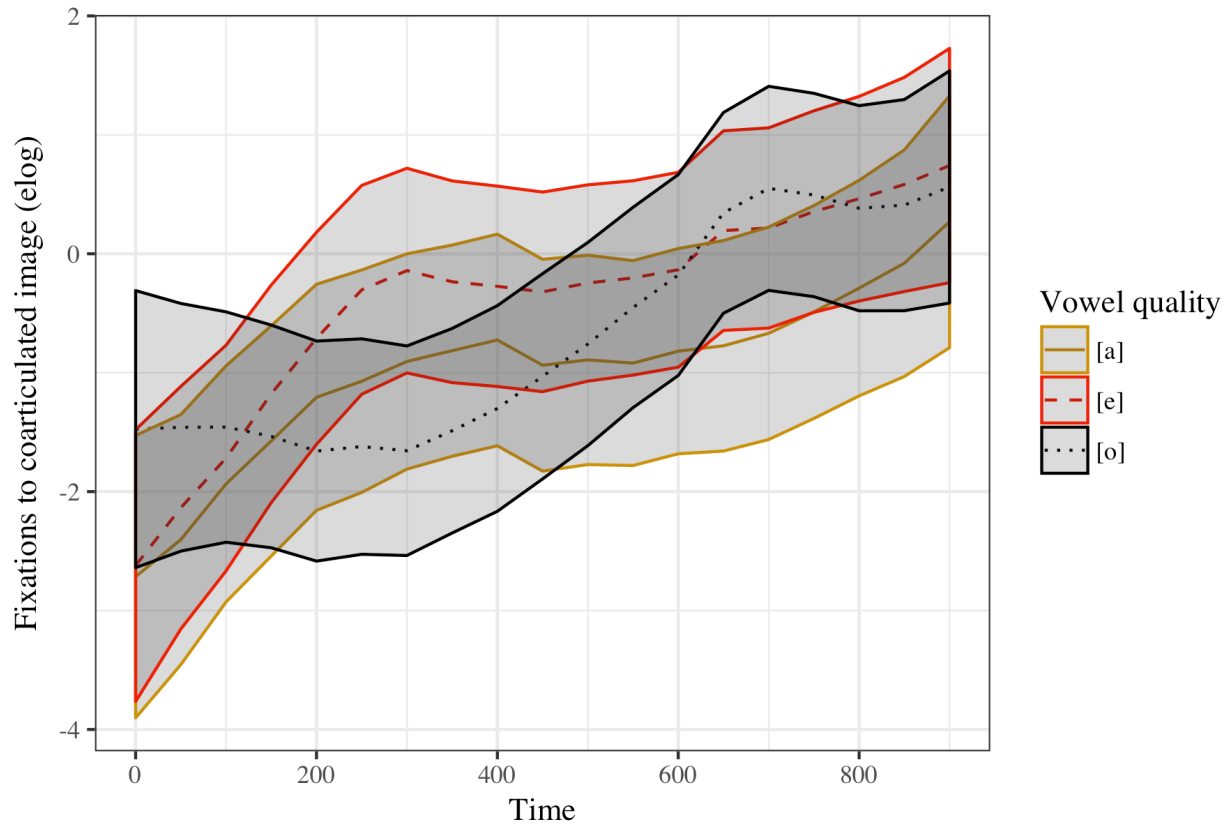


Figure 3.24. Predicted data by VOWEL QUALITY through TIME – Exp. 1b elog %fix to CVN, CVN choices

Similarly to the other models above, the three-way interaction between %NASDUR and VOWEL QUALITY through TIME turned out significant in the GAMM. In Figure 3.25, the fixation patterns seem highly complex, potentially obscuring their interpretation. For example, in the 100N %NasDur, fixations to the CVN image are stable followed by a steep increase for [a], begin with a steep increase followed by a stabilization for [e], and do not seem as differentiated through time for [o] items. This is further complicated by the fact that this model only analyzed 198 trials, therefore strong interpretations are avoided, and further research will be necessary to precisely assess the influence of vowel quality on processing of nasal and coarticulated vowels.

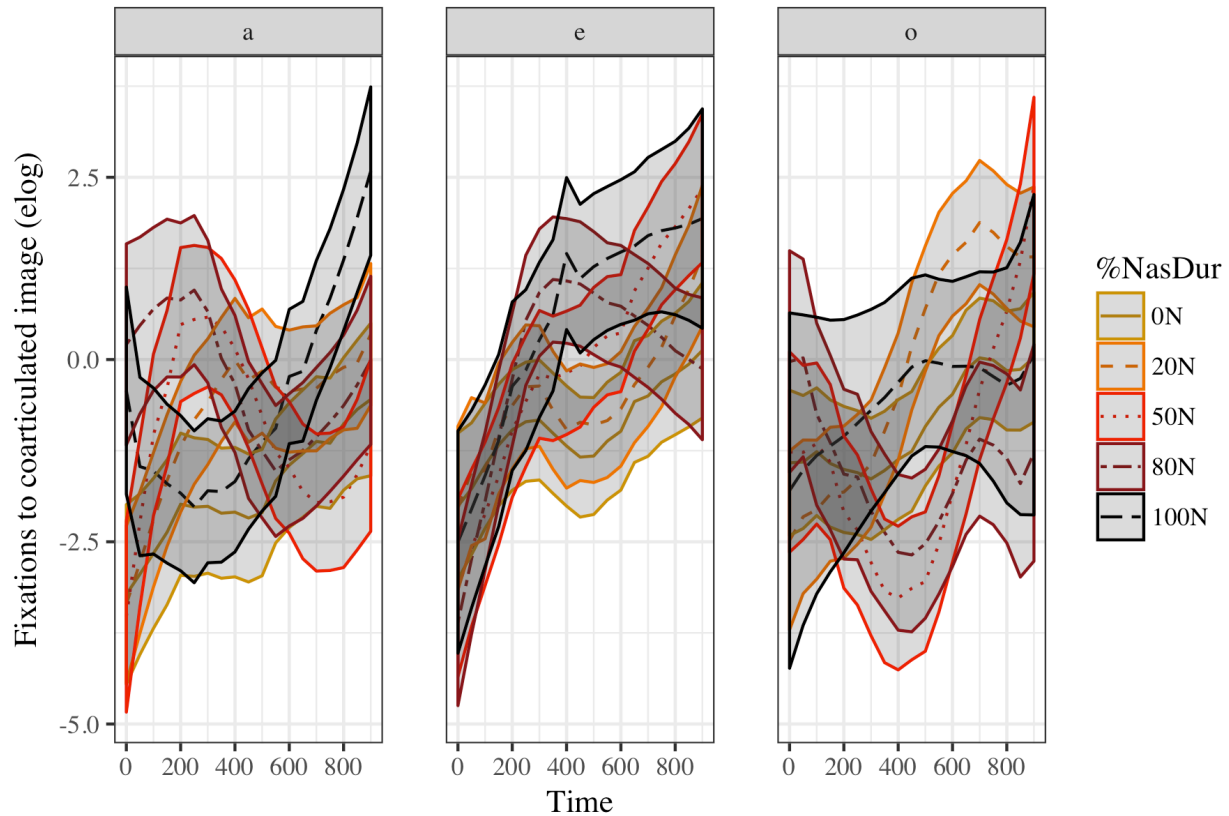


Figure 3.25. Predicted data by %NASDUR and VOWEL QUALITY through TIME – Exp. 1b elog %fix to CVN, CVN choices

3.2.3 Additional results and discussion

The results of Experiment 1b in Section 3.1.2 show that the duration of nasalization on the vowel, i.e. the %NASDUR factor, had a significant impact on Canadian French (CF) listeners' responses (image choices, RTs and eye movements) in the spoken word recognition task. The number of times participants chose the phonological nasal ($C\tilde{V}$), coarticulated (CVN), or oral (CVC) word, which corresponds to the chosen image measure, reaction times (RTs), and proportions of fixations were all influenced by the duration of nasalization on the vowel. For example, results show that when a vowel is nasalized for a longer period (e.g. 50N, 80N and 100N), participants overwhelmingly identified the stimuli as instances of the $C\tilde{V}$ word (between

92% and 95% of the time). However, when participants heard a word with a vowel that was nasalized for a shorter period (i.e. 20N), the rate of identification as C $\tilde{V}$ was slightly lower (82%). They also identified stimuli in this condition as instances of a CVN word in approximately 14% of cases (as opposed to less than 7% in the 50N, 80N and 100N %NASDUR). Analysis of word choices suggested that the probability of choosing the C $\tilde{V}$ image in the 0N and 20N %NasDur conditions was significantly different from the probability in the 50N %NasDur. However, the probabilities in the 80N and 100N %NasDur were not different from 50N, suggesting a categorical pattern of recognition, which is likely due to a ceiling effect once 50% nasalization duration on the vowel is reached.

Interestingly, the stimuli in the 0N %NASDUR were identified as instance of C $\tilde{V}$ words at a rate of approximately 34%. Note that these stimuli did not fully correspond to one of any of the critical images displayed on the screen (e.g. the auditory stimulus [pɛ̃] did not correspond to the images for *pain* /pɛ̃/ ‘bread’, *peigne* /pɛɲ/ ‘comb’, or *pêche* /pɛʃ/ ‘peach’), and that the vowel was not nasalized at all in this case. It was spliced from the CVC (e.g. *pêche* /pɛʃ/ ‘peach’) word, with no nasalization, and formant transitions that led to the place of articulation of the consonant in the oral (CVC) word. This can explain the identification rate of approximately 44% of these stimuli as instances of the oral CVC category. Participants appeared sensitive enough to vowel transitions towards vowel offset, which has been demonstrated using a similar paradigm in Dahan et al. (2001b). Even though it does not directly concern vowel nasalization, this result supports the inclusion of fine-grained phonetic information into word representations, and the hypothesis that speakers are able to retrieve this information when stimuli are ambiguous (in this case, they did not correspond to any of the presented images, therefore participants picked the best option for what they heard).

A significant effect of %NASDUR was also observed consistently for RTs and eye movements in the statistical analysis of RTs and proportions of fixations to the nasal (C $\tilde{V}$) image (C $\tilde{V}$ word choices). This is similar in Experiment 1a, in that RTs were shorter when identification was consistent with canonical realizations of C $\tilde{V}$ s and CVNs. Throughout the experiment, participants' RTs were longer for vowels that were nasalized for a shorter period (0N and 20N), than when they were nasalized for a longer period (50N, 80N, and 100N). For the latter %NASDUR values, RTs were similar in both the raw data (Figure 3.17) and predicted data by the GAMM analysis (Figure 3.18). For the proportions of fixations¹¹, a similar pattern emerged concerning the %NASDUR factor. When the vowel was not nasalized (0N) or nasalized for a short period (20N), participants did not fixate the nasal (C $\tilde{V}$) word as much through TIME as in the other three %NASDUR (50N, 80N, 100N). Again, proportions of fixations to the nasal (C $\tilde{V}$) image was highly similar for the latter three %NASDUR (see Figure 3.20**Erreur ! Nous n'avons pas trouvé la source du renvoi.** and Figure 3.21). These results suggest that when a vowel is nasalized for 50% of its duration or more, processing is similar, and variability past this threshold does not seem to impact recognition as much as when it is nasalized for a shorter period. However, when a vowel is not nasalized (0N) or nasalized for a short period only (20N), processing seems more demanding. This may be due to the fact that not- or short-nasalized vowels that are not followed by a nasal consonant after the word are not consistent with the production of neither phonological nasal ($\tilde{V}$) or coarticulated nasalized (VN) vowels in CF. Indeed, Desmeules-Trudel and Brunelle (2018) showed that in production, vowels that are nasalized for 20% or less of their duration are always followed by a nasal consonant (CVN

¹¹ Note that the current discussion focuses on the analyses of fixations to the C $\tilde{V}$ image and only on trials in which the nasal (C $\tilde{V}$) image was chosen, since they represent 89% of the trials in Experiment 1b. The discussion of the CVN trials would not enable me to draw conclusions based on such a small number of trials.

word). Therefore, a vowel that is nasalized for a short period and not followed by a nasal consonant does not correspond to either a coarticulated (CVN) vowel, nor a phonological nasal (C $\tilde{V}$) vowel. This likely makes recognition of the auditory word harder than when the physical characteristics correspond to the realization of one of the categories, i.e. when the vowel is nasalized for a longer period and thus corresponds to the realization of a phonological nasal (C $\tilde{V}$) vowel, as reflected in RT values.

Experiment 1b was designed to assess the gradient recognition hypothesis of phonetic cues, when stimuli do not contain conflicting phonetic cues. As mentioned above, removing the nasal consonant at the end of the vowel for some stimuli disambiguated their phonological status (especially for 50N, 80N and 100N). However, note that this modification made the 0N stimuli correspond to none of the presented word choices, therefore making these stimuli a poor realization of all the presented image choices. The 20N %NASDUR stimuli also did not fully correspond to any of the proposed image choices, except perhaps as a poor realization of the C $\tilde{V}$ word. This can explain the slightly lower (but well above chance) rate of identification of stimuli in the 20N condition as C $\tilde{V}$ words. Keeping in mind the differences across conditions, the predictions of the gradient recognition could be further tested using similar stimuli. For example, intermediate steps (e.g. additional %NASDUR values in 5% increments, such as 25N, 30N, 35N, etc.) could lead to gradient patterns of word recognition and phonetic integration in Experiment 1b, especially surrounding the perceptual boundary between 20% and 50% nasalization duration – more research will be necessary to test this hypothesis. However, when the identity of the vowel is disambiguated (once the 50% nasalization threshold is reached), the identification of the stimuli seems more categorical and yields a ceiling effect. Participants consistently identify these stimuli as phonological nasal (C $\tilde{V}$) words, and the time course of fixations seems similar across conditions; there seems to be a ceiling effect in the 50N or more %NASDUR values.

In addition to the factors that enabled me to document the questions of the influence of nasalization duration on word recognition and phonetic integration, other factors came out significant for the statistical analyses of proportions of fixations to the nasal ($C\tilde{V}$) image. For example, a significant nonlinear interaction between TIME and TRIAL NUMBER was found in the GAMM (Figure A.21 in Appendix 3). Although hard to interpret from the visual inspection of the figure, the significance of the smooth in Table 3.15 suggests that proportions of fixations to the nasal ($C\tilde{V}$) image through time differed across the duration of the experiment, and that they became higher towards the end of the experiment than at the beginning (see darker orange region between trials 100 to 150 on the y-axis, and after 600 ms on the x-axis). Even though this does not seem highly robust, it is reasonable to think that participants tended to anticipate the $C\tilde{V}$ stimuli over the course of the experiment.

3.3 GENERAL DISCUSSION

Two experiments showed that the duration of vowel nasalization has a major impact on the recognition of words containing phonological nasal ($C\tilde{V}$) and coarticulatory nasalized (CVN) vowels in Canadian French. This influence was predicted based on the production of the contrast in Canadian French (Desmeules-Trudel & Brunelle, 2018), and on previous results obtained in English by Beddor et al. (2013). For example, Desmeules-Trudel and Brunelle (2018) found that phonological nasal vowels are produced significantly differently from coarticulated nasalized vowels in Canadian French throughout their duration. We also found that nasalization of phonological nasal vowels ($C\tilde{V}$) starts between the vowel onsets and 50% of their duration, and nasalization of coarticulated nasalized vowels (CVN) starts later. Participants recognized vowels for which nasalization started between onset and 50% of their duration as $C\tilde{V}$ more often than

when it started later in the vowel, especially in Experiment 1b. However, when there were conflicting cues in the word (i.e. a vowel that was nasalized early followed by a nasal coda), especially in Experiment 1a, participants chose the CVN image most of the time. Importantly, an increase in duration of nasalization yielded more $C\tilde{V}$ responses in both experiments. The results are also consistent with Beddor et al.'s (2013) results that listeners were sensitive to variations in nasalization timing. I draw a parallel between Beddor et al.'s (2013) participants who were more efficient at anticipating an upcoming nasal consonant when the vowel was nasalized early, and the participants to the current experiments who tended to recognize words that contained a phonological nasal vowel ($C\tilde{V}$) in a greater proportion when the stimuli were nasalized early.

In addition to the question of the influence of nasalization timing on recognition, by varying the duration of nasalization on the vowel, I found support in favour of gradient recognition of time-dependent phonetic information (McMurray et al., 2002; 2008), especially when the stimuli did not correspond to the canonical realization of a phonological nasal or a coarticulatory nasalized vowels. Indeed, in Experiment 1a, the gradient pattern of recognition emerged when a nasalized vowel was followed by a nasal consonantal appendix, and when participants interpreted these stimuli as instances of phonological nasal ($C\tilde{V}$) vowels. Consequently, it seems like gradient recognition emerged for words that contained ambiguous cues or conflicting phonetic cues regarding their identity. On the other hand, when stimuli did not contain conflicting cues (when vowel nasalization duration was short and followed by a short nasal consonant in 1a, or when nasalization duration was longer and not followed by a consonant in 1b) more categorical patterns of recognition emerged. This is an important finding, that listeners are able to use a great amount of fine-grained phonetic information, but only when stimuli are ambiguous and it is necessary to disambiguate stimuli.

In addition to providing support for the gradient integration hypothesis when stimuli contain conflicting phonetic cues, my results provide support for the storage of fine-grained phonetic cues in phonological representations (McMurray et al., 2002) in general, as listeners were able to consider short variations in the signal to categorize the stimuli. The fact that listeners are able to use fine-grained phonetics for higher-order language processing has been found in English using a variety of phonetic cues, such as VOT and length of vowel transitions (McMurray et al., 2008), direction of vowel transitions (Dahan et al., 2001b), and vowel nasalization (Beddor et al., 2013; Zamuner et al., 2016). One of the main additions of my study to the literature on coarticulation processing lies in the fact that listeners are able to use these cues, but do not do so when the task is simple or when coarse cues are made available. On the other hand, when the task demands are greater or when the stimuli present conflicting cues, listeners can rely on more fine-grained strategies for interpreting spoken words.

Finally, as one of the basic motivations for this dissertation was to extend the results on phonetic integration with native English listeners to other populations, I will present two additional experiments with second language listeners of CF in the following chapters. This will enable me to provide more details about the structure of phonological representations and the importance of phonetic details for word recognition.

4 Second language speech perception and word recognition

In the previous chapters of this dissertation, I have shown that native (L1) listeners of Canadian French are able to use fine-grained phonetic details in a gradient manner during word recognition, under certain stimuli conditions. Evidence from the two eye tracking experiments presented in Chapter 3 demonstrated both gradient and categorical (i.e. ceiling) patterns of word choices and eye movements, and reaction times showed differences in processing load depending on the congruence of the phonetic cues with the participants' identification of the auditory stimuli. These data support a gradient model of speech perception and word recognition (McMurray et al., 2002) when additional low-level information is necessary to disambiguate or compensate for conflicting cues, and that coarticulatory information can be stored in phonological representations (although not necessarily always used for recognition).

It has repeatedly been shown that second language (L2) speech perception is different from L1 perception (Best, 1995; Best & Tyler, 2007; Flege, 1995). I am thus interested to see if and how phonetic details, specifically vowel nasalization duration, are considered in L2 processing of Canadian French during spoken word recognition. I will focus on listeners whose native language is English, because English has coarticulatory vowel nasalization but no phonological nasal vowels in its inventory. For learning French as a L2, English listeners have to learn how to phonologically discriminate between words that have coarticulatorily nasalized vowels and phonologically nasal vowels. In other words, they must learn a new category (i.e. phonological nasal vowels) based on a characteristic that exists in their native language (i.e. vowel nasalization), but with a different phonological status (coarticulatory vs. phonological nasalization).

The current chapter presents a review of the literature on non-native and L2 speech perception, L2 lexical access and recognition, as well as additional factors that pertain to individual speaker-listeners that can influence L2 phono-lexical representations. More precisely, I will present the *Perceptual Assimilation Model* of foreign speech perception (Best, 1995), its extension for L2 learners (Best & Tyler, 2007), and Flege's (1995) *Speech Learning Model*. It is logical that in order to recognize spoken words in an L2, which is the object of investigation of the experiments in Chapter 5, listeners must be able to discriminate sounds that are not present in their native language. It is exactly the purpose of the speech perception models enumerated here. In the current chapter, I will thus limit my presentation to the models and experimental evidence to support them. I will also present a brief comparison between the phonetic-phonological properties of vowel nasalization in (North American) English and Canadian French, to allow me to establish how native English listeners might transfer, if they do, their L1 knowledge into L2 processing of Canadian French vowel nasalization.

This literature review will lead to Chapter 5, which presents two eye tracking experiments that were conducted with L2 listeners of Canadian French (native speakers of Canadian English). These experiments examine the question of phonetic integration in an L2 as well as other factors on the perception of an L2, using the same methods and paradigm as for L1 speakers in Experiment 1a and 1b. As mentioned above, one of the main reasons for investigating English-native learners of French lies in the fact that Canadian French has a phonological contrast between nasal ($\tilde{V}$) vowels and oral vowels (which can also be coarticulatorily nasalized when followed by a nasal consonant), but that English does not have contrastive nasal vowels in its phonological inventory. In English, nasalized vowels are the result of a nasal coarticulation rule and are not “phonologically specified” for nasality (Beddor, 2009; Cohn, 1990). In order to discriminate between two French lexical items that differ based on the nasality of the vowel, such

as *pain* [pɛ̃] ‘bread’ ~ *peigne* [pɛɲ] ‘comb’, L2 listeners (L1 English) have to learn that nasalized vowels can be phonological in French. One of the main cues to the phonological French distinction is the duration of nasalization, which is time-dependent in the speech signal. L2 listeners thus have to be sensitive to variations in nasalization duration to accurately recognize words that contain phonological nasal vs. coarticulatory nasalized vowels.

4.1 NAÏVE SPEECH PERCEPTION

Naïve (foreign) speech perception is defined as the way listeners hear the sounds of a foreign language, without having previous knowledge of the target language (Best, 1995; Best & Tyler, 2007). This represents the initial state for listeners’ perceptual abilities for foreign speech sounds, in other words, before any learning has occurred. This initial state has been shown to be a good predictor of the difficulties encountered when learning an L2 phonological system (Best & Tyler, 2007; Escudero, 2005). Consequently, before learning the phonetic and phonological systems in their L2, it is assumed that foreign listeners will transfer the knowledge on their L1 onto L2 categories (Major, 2008), meaning that they will perceive foreign sounds according to their L1.

4.1.1 *The Perceptual Assimilation Model*

One of the most influential models of naïve speech perception is the *Perceptual Assimilation Model* (PAM; Best, 1995). PAM is concerned with how foreign sounds are perceptually *assimilated* for identification. Here, assimilation is a type of categorization of speech sounds: when hearing a speech sound, the listener will assign it a category according to the existing ones in their native phonological system. Assimilation can be seen as analogous to the

language transfer process. For example, listeners can assimilate a foreign sound to a native (L1) category, to a speech sound that falls within the phonological space but not into one specific L1 category, or to a non-speech category. The way individual sounds are assimilated is then expected to yield different *discrimination* patterns when investigating contrasts (that necessarily involve two sounds; Best, 1995):

- Two-category discrimination: Each sound is assigned to a different L1 category, and discrimination is expected to be excellent;
- Category-goodness difference: The two sounds are assimilated to the same L1 category, but differ in their discrepancy to the canonical realization in the L1 (e.g. one is a good exemplar, the other is a poor exemplar), and discrimination is expected to be moderate to very good;
- Single-category assimilation: Both sounds are assimilated to the same L1 category and are equally deviant from the canonical L1 realization, and discrimination is expected to be poor;
- Both uncategorizable: Both speech sounds fall outside of the L1 categories (but *within* the phonological space; e.g. a non-existing vowel in the L1), and discrimination is expected to greatly vary across contrasts;
- Uncategorized-Categorized discrimination: One sound is not assimilated to any L1 category and the other is assimilated to an L1 category, and discrimination is expected to be excellent;
- Nonassimilable: Both sounds fall outside of the listener's L1 categories and *outside* of the phonological space (e.g. click consonants as perceived by a native speaker of English), and discrimination is expected to be good to very good.

All of these patterns have been extensively tested, and predictions have been repeatedly borne out, especially using consonantal contrasts with naïve or inexperienced listeners in the L2 (Best & McRoberts, 2003; Best, McRoberts, & Goodell, 2001; Guion, Flege, Akahane-Yamada, & Pruitt, 2000). However, some studies have also investigated L2 vowel assimilation and vowel discrimination patterns (Escudero & Chládková, 2010; Escudero, Sisinni, & Grimaldi, 2014; Escudero & Williams, 2011; Escudero & Vasiliev, 2011; Gilinchinskaya & Strange, 2010; Levy, 2009; Levy & Strange, 2008; Tyler, Best, Faber, & Levitt, 2014). As the current work focuses on the perception, learning and recognition of a vocalic contrast (i.e. nasal vs. non-nasal vowels in L2 French), the remainder of this discussion will focus on studies of vowel discrimination.

4.1.2 Experimental evidence in favour of PAM classification

Based on the predictions formulated by PAM, Tyler et al. (2014) assessed the assimilation and discrimination of foreign vowels from different languages by American English listeners. In a first task, which focused on assimilation of single vowels, listeners had to say to which English vowel category corresponded each experimental foreign sound (i.e. vowels from Norwegian, Thai and French). Tyler et al. (2014) thus were able to assess if listeners consistently assimilated the vowels to an English category (i.e. categorizable vowel), or if they variably assimilated them to more than one category (i.e. uncategorizable vowel). In a second task (i.e. AXB test), they investigated listeners discrimination performance on vowel pairs and were able to link the single-vowel assimilation and the vowel-pairs discrimination patterns. In general, listeners performed well on the stimuli that were discriminated according to the two-category (e.g. [dø]~[dœ]), uncategorizable-categorizable (e.g. [bu]~[bʏ]), and uncategorizable-uncategorizable (e.g.

[bo]~[bõ]¹²) patterns. This was expected based on PAM predictions (see Section 4.1.1 above). However, both (Norwegian) [i] and [y] vowels were assimilated to the (English) listeners [i] category in the first task. Consequently, the [ki]~[ky] pair, which was considered a single-category contrast, and was poorly discriminated in the second task. This result was also expected based on PAM predictions.

One of the factors that can explain the difficulties in naïve non-native speech perception is the acoustic proximity of the members of the contrast. Note that PAM is based on differences and similarities between “constellations of articulatory gestures”, and not on differences or similarities on the acoustic signal that listeners are presented with. However, the focus of the current work is not to be placed on the debate concerning the articulatory vs. acoustic units of speech perception (i.e. Liberman and Mattingly’s (1985) motor theory of speech perception and Ohala’s (1996) acoustic theory of speech perception). I will consider acoustic differences as the source of the observed differences for listeners’ discrimination abilities, as acoustic proximity can be easily measured for vowels (e.g. Euclidian distance within a phonetic space, differences in durational properties of vowels).

As an example of how cross-linguistic acoustic proximity can explain foreign sound classification, Escudero and Vasiliev (2011) investigated the differences in perception of Canadian French and Canadian English [ɛ]~[æ] contrast by native Peruvian Spanish speakers. They found that the acoustic proximity of each foreign vowel to the native Spanish categories predicted the perceptual assimilation patterns achieved by listeners. In Figure 4.1, reproduced from Escudero and Vasiliev (2011, p. EL278), it is shown that the Canadian French vowels (black font [ɛ] and [æ]) are acoustically closer (in the F1/F2 dimensions) to two Spanish

¹² Although it may seem surprising that the French [bo] vowel was assimilated as uncategorizable in English, the results of the identification experiment suggest that this French vowel was categorized as English [ow] 44% of the time, and as English [u] 38% of the time.

categories ([e] and [a] within ellipses), as opposed to Canadian English vowels (grey font [ɛ] and [æ]), which are closer to a single Spanish [a] category. Spanish listeners overwhelmingly categorized French [ɛ] as Spanish [e], and French [æ] as Spanish [a], while they tended to categorize both English vowels as Spanish [a] overall. Using linear discriminant analysis, the authors found that the acoustic proximity (on F1, F2 and F3) of the French and English vowels significantly predicted the perceptual assimilation results. This suggests that acoustic proximity of the languages' vowels is a major predictor of listeners' performance on a foreign sound assimilation task. As the current work focuses on the perception of vowel nasalization in L2 Canadian French by native speakers of Canadian English, a description of the acoustic output of vowel nasalization in English is provided in Section 4.5 along with a description of the similarities and differences to Canadian French vowel nasalization patterns.

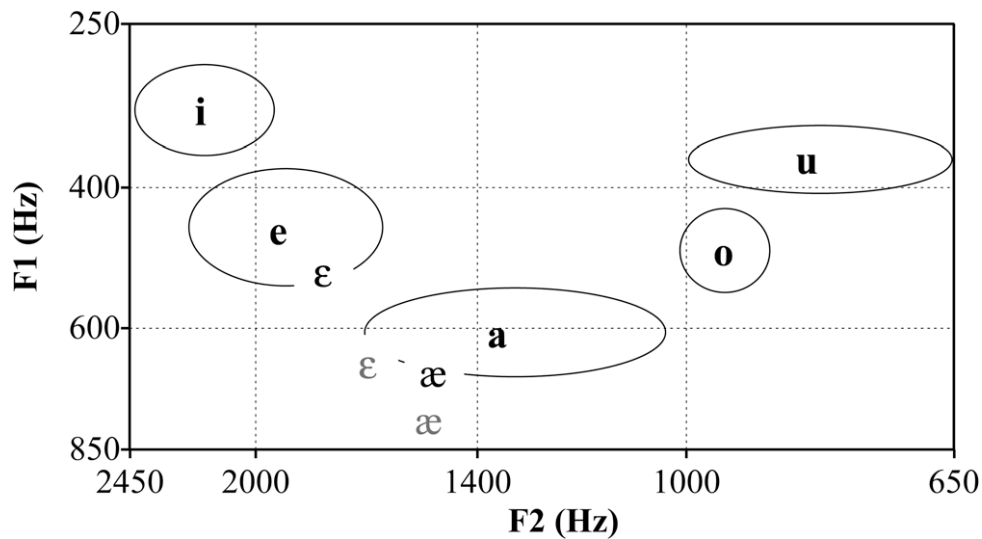


Figure 4.1. Illustration of the acoustic proximity (F1/F2) of Canadian French (black-font ϵ and æ) and Canadian English (grey-font) $[\epsilon]$ - $[\text{æ}]$ vowels compared to Peruvian Spanish $[e]$ - $[a]$ vowels, reproduced from Escudero and Vasiliev (2011, p. EL278). Ellipses correspond to the dispersion of Spanish vowels in the F1/F2 phonetic space (2 standard deviations to the mean).

4.2 SECOND LANGUAGE SPEECH PERCEPTION

Even though PAM makes interesting predictions on how foreign sounds will be perceived, it models non-native speech perception before language learning has occurred. Other models, such as Flege's (1995) *Speech Learning Model* (SLM) and Best and Tyler's (2007) PAM extension to L2 speech perception (PAM-L2), have thus been developed to account for patterns within L2 learners. In these models, it is assumed that learning sounds in an L2 eliminates the need for listeners to transfer L1 knowledge into their L2. Consequently, more advanced L2 listeners will perceive L2 sounds in a more "native-like" manner, while beginners will tend to transfer (Major, 2008) their L1 categories onto L2 sounds for identifying, discriminating and interpreting L2 categories (Flege, 1995).

4.2.1 *The Speech Learning Model*

In Flege's (1995) SLM, one of the most influential models of L2 speech perception and production, it is assumed that errors in production and accented speech are rooted in perception (as opposed to a lack of phonetic input, insufficient motivation, etc.). That is, *perceptual* distortions of the L2 input sounds, due to the influence of the L1 system, impact L2 representations and are thus at the basis of accented speech in the L2. Consequently, documenting the perceptual patterns of L2 learners and listeners will enable us to predict the difficulties in the production of foreign speech sounds, according to Flege (1995).

The SLM is based on a number of postulates for the perception and production of foreign speech sounds. For example, Flege (1995) postulates that the category formation mechanisms remain intact over the life span, and that the pre-existing categories evolve throughout life. Furthermore, L1 and L2 phonetic categories coexist in a common phonological space. Given this,

it is hypothesized that L1 and L2 sounds are perceptually related to one another at an allophonic-positional level, suggesting that their representations include some amount of phonetic detail. It is also hypothesized that new phonetic categories develop if listeners are able to discern at least one phonetic difference from a pre-existing L1 category, and that greater phonetic differences yield to easier discrimination and thus easier category formation. Finally, SLM predicts that category formation may be blocked by equivalence classification, which is similar to single-category assimilation in PAM. In this case, two L2 sounds are attributed to the same phonological category and it will be more difficult for L2 speakers to produce them.

4.2.2 *The Perceptual Assimilation Model for second language perception*

Flege's (1995) SLM was mostly designed to explain production errors and accented speech phenomena. Thus, based on the SLM postulates and hypotheses as well as on PAM predictions on naïve perception, Best and Tyler (2007) formulated an extended PAM model to account for speech perception and phonetic learning in the L2. Firstly, it is important to note that in PAM-L2, similarly to PAM, discrimination abilities of L2 learners depend on the phonetics of the contrast, and not necessarily on more abstract phonemic information. Furthermore, it is assumed that discrimination abilities differ across L1s and L2s, and across dialects (Brannen, 2002). It is thus possible that a great deal of variability may be observed in the study of L2 perceptual abilities for *linguistic* reasons as well as listener-specific reasons (see Section 4.4).

Concretely, PAM-L2 formulates four scenarios that are likely to occur during perceptual learning (summarized in Table 4.1). Firstly (Scenario A in Table 4.1), it is possible that L2 learners assimilate one member of a contrast to an L1 category, and that the other member is assimilated to another native or non-native category. In this case, no learning is necessary for the

first member of the contrast, and discrimination is expected to be very good (two-category assimilation or uncategorized-categorized discrimination; see Section 4.1.1). Secondly (B), L2 listeners may perceive both members of the contrast as instances of the same L1 category, but one can be perceived as more deviant from the canonical realization. Here, discrimination is expected to be good (same as category-goodness differences; see Section 4.1.1), perhaps not as good as when the two members of the contrast have been categorized as instances of two distinct phonological categories. Perceptual learning will also be fairly easy, as phonetic differences between each member of the pair are perceived. Thirdly (C), it is possible that the two sounds of the L2 contrasts would be perceived as equally poor or equally good instances of the same L1 category. In this case, perceptual learning is expected to be difficult, because they would have to learn the category of at least one of the two members of the contrast before being able to discriminate them successfully. Also, discrimination of this contrast is expected to be poor (single-category assimilation; see Section 4.1.1). Finally (D), the last scenario that is considered by Best and Tyler (2007) is when there is no assimilation of the L2 contrast to any of the L1 categories. In this case, depending on the contrast, discrimination is expected to be fairly good (both uncategorizable sounds; see Section 4.1.1), and learning is expected to be fairly easy. Table 4.1 summarizes the four sets of predictions formulated by PAM-L2 for discrimination and learning abilities of L2 learners.

Table 4.1

Summary of the PAM-L2 predictions for discrimination and perceptual learning abilities in an L2

	Scenario	Effect on discrimination	Effect on learning
A.	Only member A assimilated to L1 category	Very good	Easy learning
B.	Member A is a good instance of L1 category, member B is a poor instance	Good	Easy learning
C.	Both members are equally good/poor instances of the L1 category	(Fairly) poor	Difficult learning
D.	No member is attributed an L1 category	(Fairly) good	(Fairly) easy

These scenarios have been tested using a variety of language pairs (Bundgaard-Nielsen, Best, & Tyler, 2011; Levy, 2009; Levy and Strange, 2008). For example, Levy and Strange (2008) compared experienced and inexperienced learners of French as an L2 (English L1) and tested their discrimination abilities of the French front-back and rounded-unrounded contrasts. In three out of four contrasts, experienced listeners performed better than inexperienced learners. The French [i]~[y] (rounding distinction; note that this is not consistent with Tyler et al.'s (2014) results with Norwegian [i]~[y]), [u]~[œ] (frontness and height distinction) and [y]~[œ] (height distinction) contrasts were well discriminated overall, and all fell within the A or D scenarios of PAM-L2. However, the [u]~[y] contrast was poorly discriminated by all listeners (especially when surrounded by alveolar consonants). Factoring in the surrounding context of the vowel, Levy and Strange suggest that both French [u] and [y] fall within the same English category for listeners (as monophthong [u] and front-rounded [y] are both poor instances of English [u]), and that learning is thus difficult for English speakers (scenario C in Table 4.1).

The experiments that are presented in Chapter 5 were designed to verify if L1 English listeners are able to discriminate between phonological nasal and coarticulatory nasalized vowels in L2 CF, but in the context of spoken words, as this is thought to more closely reflect the real-life phenomena of L2 language learning. Since English natively has coarticulatory nasalized vowels, L2 listeners of CF are expected to behave according to the B scenario of Table 4.1. However, the spoken words context could confound the results by increasing the task demands, as opposed to a simpler syllable-discrimination task.

4.3 LEXICAL RECOGNITION IN A SECOND LANGUAGE

The models of L2 perception that were presented above were mainly developed based on discrimination and identification tests that involved isolated syllables or nonwords. This means that PAM, SLM and PAM-L2 model perceptual abilities of L2 listeners without necessarily considering the need to recognize words and associate them with meaning when processing language (as well as other phonological factors such as phonotactics or prosody). Indeed, most L2 learners' goal is to understand the meaning that is conveyed while listening to the target language, and a large part of that meaning is contained within lexical items.

On the one hand, psycholinguistic models of spoken word recognition, such as TRACE (McClelland & Elman, 1986), SHORTLIST (Norris, 1994) or PARSYN (Luce et al., 2000) do not formulate specific predictions or accounts of spoken word recognition when more than one language is involved. On the other hand, while the L2 lexical access BIA+ model (Dijkstra & van Heuven, 2002) postulates sublexical- and lexical-phonological levels, its focus is mainly on the recognition of semantic properties of lexical items and how meaning is linked between languages in the bilingual mind. Importantly for the current dissertation, it does not make specific claims on

how subphonemic information, such as coarticulation and fine-grained phonetic details, is integrated during L2 spoken word recognition: BIA+ is essentially a model of orthographic word recognition. Note that the goal of the current work is not to criticize models of word recognition, neither to propose substantial modifications to account for L2 *spoken* word recognition. However, as demonstrated by the lack of modeling of such a process, it is evident that little research has been completed on phonetic integration during L2 spoken word recognition. The current work aims to overcome this gap in research by documenting how similar or different L2 phonetic integration of French vowel nasalization by native English listeners is from native French listeners.

Weber and Broersma (2013) review some of the main findings on L2 spoken word recognition and argue that L2 listeners cannot completely separate their two lexical systems (i.e. integrated lexicon; Sabourin, Leclerc, Burkholder, & Brien, 2014), including the phonological word forms. In addition, if a listener is not able to separate the L2 phonology from their L1 phonology, they might have difficulties in discriminating two L2 words if there is phonetic confusion (analogously to non-native and L2 speech perception processes reviewed in Sections 4.1.1 and 4.2.2). To illustrate, Weber and Cutler (2004) showed that Dutch speakers listening to (British) English tended to fixate competitor items more when the target English vowel could be confused with a Dutch vowel than when the target vowel was not confusable. In their experiment, when listeners heard the spoken word *panda*, they fixated more the image of a *pencil*, as English [æ] and [ɛ] are often confused by Dutch speakers. On the other hand, when they heard *bottle* and were presented with a competitor *beetle* image, they did not fixate the *beetle*. This is due to the fact that English [ɑ] and [i] are not confused by Dutch listeners. This suggests that the L1 Dutch phonology, that does not have a contrast between [ɛ] and [æ] and explains why these two vowels are confused, had an influence on the L2 eye movement patterns (this is despite the participants

high proficiency levels in the L2). Overall, then, it seems like the phonological structure of the L1 has a major influence on the L2 word recognition process, and that SLM and PAM-L2 can be extended to word recognition patterns, in addition to speech perception.

Other factors also impact the recognition of spoken words in an L2. Here I focus on word frequency, because it is one of the key factors to spoken word recognition in the L1 (Dahan et al., 2001a) and has also been shown to impact L2 recognition (Duyck, Vanderlест, Desmet, & Hartsuiker, 2008). It is also very likely that L2 listeners will tend to know and master frequent words to a greater extent than unfrequent words (Koirala, 2015), which is why this factor is assessed in the analyses below. For example, Duyck et al. (2008) used a lexical decision task on visual words with L2 English (native Belgian Dutch) listeners, and found that frequency effects were larger in the L2 than in the L1. On average, listeners were 103 ms faster at deciding if high-frequency words of English were real words than low-frequency words, compared to 46 ms in Dutch. Overall, though, they were generally slower to respond in the L2 (between 599 and 702 ms on average) than in the L1 (between 527 and 573 ms on average). These results suggest that frequency effects can be found in the L2. However, note that “L2 word frequency” might be different from “L1 word frequency”, as L2 listeners are most likely not exposed to the same amount (or type) of input in the target language as L1 speakers, and because individual characteristics (i.e. in terms of language input, proficiency, age of acquisition, manner of acquisition) can be highly variable within L1 speakers. All of these factors may have an influence on the magnitude of frequency effects in an L1. Still, the relative frequency of words should be similar across languages, given sufficient naturalistic exposure to the target language.

Finally, research suggests that interlanguage, phonological, and lexical influences on spoken word recognition in an L2 can be explained by some general activation-inhibition processes. For example, as the number of phonologically-different lexical items is greater in the

bilingual lexicon than in the monolingual lexicon, the number of potentially activated candidates is greater in bilinguals than monolinguals (Weber & Broersma, 2013). More resources are thus necessary to inhibit all the non-pertinent candidates, and processing of L2 words becomes more costly. Similarly, frequency effects are expected to vary based on the input that listeners receive when learning their L2, but higher-frequency words tend to be activated and recognized faster than lower-frequency ones (Forster & Chambers, 1973). L2 listeners are thus expected to display frequency effects when recognizing words, analogously to L1 listeners (see Sections 3.1.3 and 3.2.3).

4.4 LANGUAGE EXPOSURE AND SECOND LANGUAGE ACQUISITION

Up to this point, I have reviewed some of the *linguistic* factors that impact speech perception and word recognition in an L2. However, as has been underlined above, many differences in L2 processing can be attributed to other factors that pertain to individual listeners. For example, researchers have hypothesized that L2 proficiency (Archila-Suerte, Zevin, Bunta, & Hernandez, 2011), the extent of L1/L2 use and the amount of L2 input or exposure (i.e. experience in the L2; Bohn & Flege, 1990), age of acquisition (AoA; Flege & Mackay, 2004), and manner of acquisition (MoA; Sabourin et al., 2014) can have an influence on L2 processing.

Although it may seem intuitive that more proficient individuals would attain more native-like performance on a speech perception task, to my knowledge, research has failed to consistently replicate L2 proficiency effects. This lack of consistency may be due, for example, to the unstable definition of L2 proficiency and on how to assess it (Tremblay, 2011), or to participants' self-reports of L2 proficiency based on potentially different criteria (Sabourin et al., 2016). While it follows that a greater amount of exposure to an L2 will contribute to proficiency

development, the effect of the L1 on L2 perception, as modeled by SLM and PAM-L2, seems persistent even for advanced bilinguals (e.g. Weber & Cutler, 2004). However, Archila-Suerte et al. (2012) found, using a syllable-similarity rating task, that highly proficient (late) L2 speakers of English (Spanish native) were better at discriminating English phonemes than low-proficiency L2 listeners, although they did not reach native-like patterns. In the current work, L2 proficiency will not be measured directly, as it is not the main variable of interest, but self-reports of spoken proficiency and general comprehension in French as an L2 will be collected using a language background questionnaire (Appendix 1). Overall, however, L2 proficiency may be a good generic term to gather information about AoA, amount of exposure, and MoA, even if it can be somewhat subjective. The remainder of this section will focus on the latter three factors and will demonstrate their impact on speech perception abilities in the L2.

Another variable that has been explored in L2 speech perception studies is the amount of input and L2 experience of L2 listeners. For example, Bohn and Flege (1990) investigated experienced and inexperienced L2 speakers of English (German natives) and found that experienced learners performed better at discriminating the English contrast between [ɛ]~[æ], which is not present in German. This suggests that experience in the L2, which may be due to earlier AoA or more extensive use of the L2 on a daily basis, has an impact on discrimination abilities. For example, as more experienced listeners were presented with more exemplars of the contrast than inexperienced listeners, it can be argued that their L2 phonological representations are more stable in the L2, and that this contributes to improving their performance on discrimination tasks.

Age of acquisition (AoA) is by far one of the variables that has been most explored in relation to L2 speech perception abilities (Flege, 1995; Wode, 1994; Yamada, 1995). For example, Flege and Mackay (2004) investigated several groups of L2 learners of Canadian

English (Italian natives) and found that early learners, who arrived young in Canada, performed better on an L2 vowel discrimination task than late learners who arrived in Canada at a later age. This suggests that, among other factors, earlier L2 learning improves perceptual discrimination abilities in an L2. In the experiments that follow, the age of acquisition of French as an L2 in the participants will be assessed (as measured by a significant – or not – amount of exposure to French at an early age) and will enable me to coarsely evaluate if early vs. late acquisition has an influence on word recognition performance.

Finally, manner of acquisition (MoA) is a factor of influence that has only recently received some attention in L2 research. For example, Sabourin et al. (2014) investigated the integration of the bilingual mental lexicon using masked (translation) priming, and showed that L2 English late bilinguals (L1 French) who were exposed to English in a natural setting tended to present a more integrated lexicon, similarly to their early-bilinguals group. On the other hand, late bilinguals who received more L2 input in a classroom setting (L2 French, L1 English) did not display reaction times that suggested an integrated lexicon for both English and French. This suggests that MoA is a significant factor of influence on the bilingual linguistic system. However, to my knowledge, no L2 speech perception study has specifically investigated this factor. For this reason, I will assess context of L2 exposure in the language background questionnaire. However, note that for all participants, MoA was conflated with AoA (early bilinguals were exposed to French in a naturalistic setting, and late bilinguals in an instructional setting), therefore the discussion of the results will be focused on AoA, which in itself includes the MoA factor.

4.5 ENGLISH VOWEL NASALIZATION AND COMPARISON WITH CANADIAN FRENCH

Before presenting the experiments that were designed to investigate spoken word recognition in L2 speakers of Canadian French, I provide a description of coarticulatory vowel nasalization in English, and briefly compare this phonetic property to French vowel nasalization (also see Section 1.2). This comparison will enable me to formulate more precise predictions on how L2 listeners of French may interpret phonological and coarticulatory vowel nasalization in French, which is contrastive, even though it is not a phonological property of their L1 (English).

Vowel nasalization is a pervasive phonetic coarticulation rule in English, as has been repeatedly demonstrated in research (Beddor, 2009; Cohn, 1990; Krakow, 1999; Solé, 1995). For example, English words like *bean* [bin] and *den* [dɛn] contain partially nasalized vowels ([bĩn] and [dɛ̃n]) due to the following nasal consonant that has an influence on the preceding vowel. However, vowel nasalization in English can also be very variable, as convincingly shown by Beddor (2009), depending on the structure of the syllable in which the vowel is found (Krakow, 1999), the spoken dialect (Solé, 1995), as well as other factors (Beddor, 2009). Importantly, vowel nasalization in English is not phonemic (Cohn, 1990) – it does not contrast lexical items.

On the other hand, French does have phonological nasal vowels in its inventory, which enables speaker-listeners to differentiate lexical items based on vowel nasality (see Section 1.2). Canadian French oral vowels can also be slightly nasalized when followed by nasal consonants (Desmeules-Trudel, 2015a; Desmeules-Trudel & Brunelle, 2018), similar to English, but with less nasalization on the vowel, i.e. the velum lowers overall later in Canadian French VN sequences than in English VN sequences when comparing results from Beddor (2009) and Desmeules-Trudel and Brunelle (2018). Consequently, in order to differentiate between French coarticulatory nasalized vowels and French phonological nasal vowels, L2 (native English) listeners have to learn how to discriminate these vowels based on differences in the duration of

nasalization on the vowel. The following chapter (5) presents two experiments that were designed to test if and how L2 listeners of Canadian French are able to discriminate these vowel categories in the context of spoken words. Note that the last Chapter (6) of this dissertation will come back on the obtained results with both L1 and L2 listeners, link them with the theoretical questions of the interface between phonetics and phonology, the definition and content of phono-lexical representations, including the notion of contrast gradience, as well as the importance of language background (L1 vs. L2) in assessing word recognition and speech perception abilities.

5 L2 perception of Canadian French vowel nasalization by native English listeners during spoken word recognition

In this chapter, I investigate how native English speakers perceive and recognize words that differ based on duration of vowel nasalization in second language (L2) Canadian French (CF). The experiments below were designed to test if fine-grained phonetic information (i.e. variations in the duration of nasalization on the vowel) is gradiently recognized (McMurray et al., 2008) during spoken word recognition, similarly to Experiments 1a and 1b in Chapter 3. Using the eye tracking method within the Visual World Paradigm (Huettig et al., 2011), the main research question is:

- Is the low-level phonetic information related to vowel nasalization integrated categorically or gradiently by L2 listeners of CF?

The current study will enable me to assess in detail how listeners process fine-grained phonetic information in the context of L2 spoken word recognition. However, the lack of literature on the topic makes it difficult to formulate a fixed set of predictions concerning the outcome of the study. The L2 listeners that participated in the experiments below were mostly university students in Ottawa and generally had a moderate to high (self-reported) level of knowledge of French (see Sections 5.1.1.1 and 5.2.1.1). However, the variability in levels of L2 proficiency within the participants is expected to have an influence on the results. Furthermore, the task involved the interaction between speech perception and word recognition abilities. The results may thus differ from PAM (Best, 1995) or PAM-L2 (Best & Tyler, 2007) predictions, as these models' purpose is mainly to account for sound discrimination abilities. Note, however, that it has been shown that English listeners are sensitive to variations in duration of nasalization in

their native language (Beddor et al., 2013), therefore it is expected that they will also be able to consider this timing information to a certain extent based on the same phonetic cue in L2 French.

Concretely, if listeners use the same strategy in their L2 as in their L1 (Beddor et al., 2013) and given that they are all moderately to highly proficient in L2 French, it is expected that eye movements and response patterns will be similar to those of native speakers. Indeed, Beddor et al. (2013) showed that English listeners were able to use vowel nasalization timing to recognize spoken words. If they are able to use vowel nasalization similarly to L1 listeners, I then expect to find gradient patterns of recognition for words that contain conflicting phonetic cues (e.g. a nasalized vowel followed by a short nasal appendix in Experiment 2a), especially when they choose C $\tilde{V}$ items. The possible predicted patterns of recognition presented in Section 3 are repeated here in Figure 5.1:

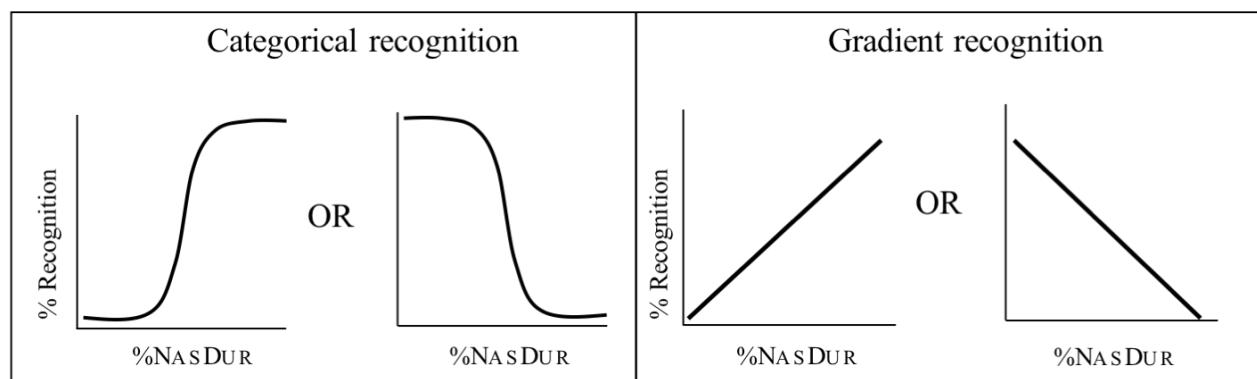


Figure 5.1. Possible predicted recognition patterns – Categorical (left) or gradient (right)

However, it could also be the case that completing the task in an L2 will increase the processing load (Clahsen & Felser, 2006), making the task harder for L2 listeners. This could lead to overall poorer performance (i.e. lower proportions of identification and fixations, longer reaction times) and perhaps more categorical patterns of recognition compared to L1 CF listeners.

The latter categorical patterns could also be due to the lack of fine-grained phonetic details in L2 phono-lexical representations, since L2 listeners do not have as much experience as L1 listeners. Reduced L2 experience will likely have an influence on the quality and quantity of information in L2 representations. The two following experiments were designed to evaluate these general predictions and to document how L2 listeners process fine-grained phonetic information during spoken word recognition.

5.1 EXPERIMENT 2A

The stimuli in Experiment 2a were the same as in Experiment 1a. Recall that they contained at times conflicting cues, i.e. the presence of a nasalized vowel and a following nasal consonant in coda position (for more details, see Sections 3.1 and 3.1.1.2). Variations of nasalization duration were also included in the stimuli as the main manipulation in the current experiments. The remainder of this section presents the methods (5.1.1), including more details on the participants (5.1.1.1), and reminders for the stimuli, procedure, and analysis (5.1.1.2 to 5.1.1.4). The presentation of the results follows in Section 5.1.2, and a discussion in 5.1.3.

5.1.1 *Methods*

5.1.1.1 *Participants*

Forty-nine L2 listeners of CF (all English-natives) were paid or received partial course credit at the University of Ottawa for their participation. Forty-four were included in the analyses (33 female, 10 male, 1 undisclosed, between 17 and 27 years old, $\bar{X} = 20.2$ years, $SD = 2.9$). An additional five participants were excluded: one listener exceeded the 40-years-old age limit (43 years old), two spoke languages that contained phonological nasal vowels (one Punjabi and one

Urdu speaker), one failed to calibrate, and one performed the task with a 16-mm lens on the eye tracker instead of the 35-mm that was used for the other participants. All listeners reported normal hearing, normal or corrected-to-normal vision, and did not report any type or history of language, hearing or speech impairment. All participants completed a background language questionnaire (see Appendix 1), and self-reported knowing French as an L2 at a low to high level of proficiency, as well as other languages. No speaker was included in the analyses that had been exposed to another language containing contrastive nasal vowels in its phonological inventory.

Seventeen listeners were early bilinguals¹³ who were exposed to L2 French in a naturalistic setting, and 27 listeners were “late” bilinguals who were exposed to French in a classroom context. The distribution of self-reported L2 French proficiency scores in the early and late bilingual groups is presented in Figure 5.2. For this score, a self-report of *speaking* proficiency in French (out of 5) and of *understanding* proficiency in French (out of 5) was asked to each participant. The *speaking* and *understanding* scores were then summed to include both components into one general L2 FRENCH PROFICIENCY score. Recall that self-reports are usually highly correlated to the actual proficiency, as measured by objective tests (Tremblay, 2011).

¹³ Let us recall that to determine if listeners were early or late French-English bilinguals, a cut-off point of 30% or more exposure to French between the ages of 0 and five years old was chosen. Research on the assessment of bilingualism level based on exposure (Thordardottir, 2011) often establishes its own criteria, making it difficult to adopt criteria based on norms in the field. I established the 30%-exposure cut-off point based on the participants’ self-reports of exposure to French, similar to Experiments 1a and 1b. The percentage of early exposure to French as an L2 is also thought to reflect the manner of acquisition (Sabourin et al., 2016) within the groups: exposure to French as an L2 before age 5 tends to be more naturalistic (non-formal instruction), while participants who were exposed to French later tended to start learning it in school, in a more formal setting.

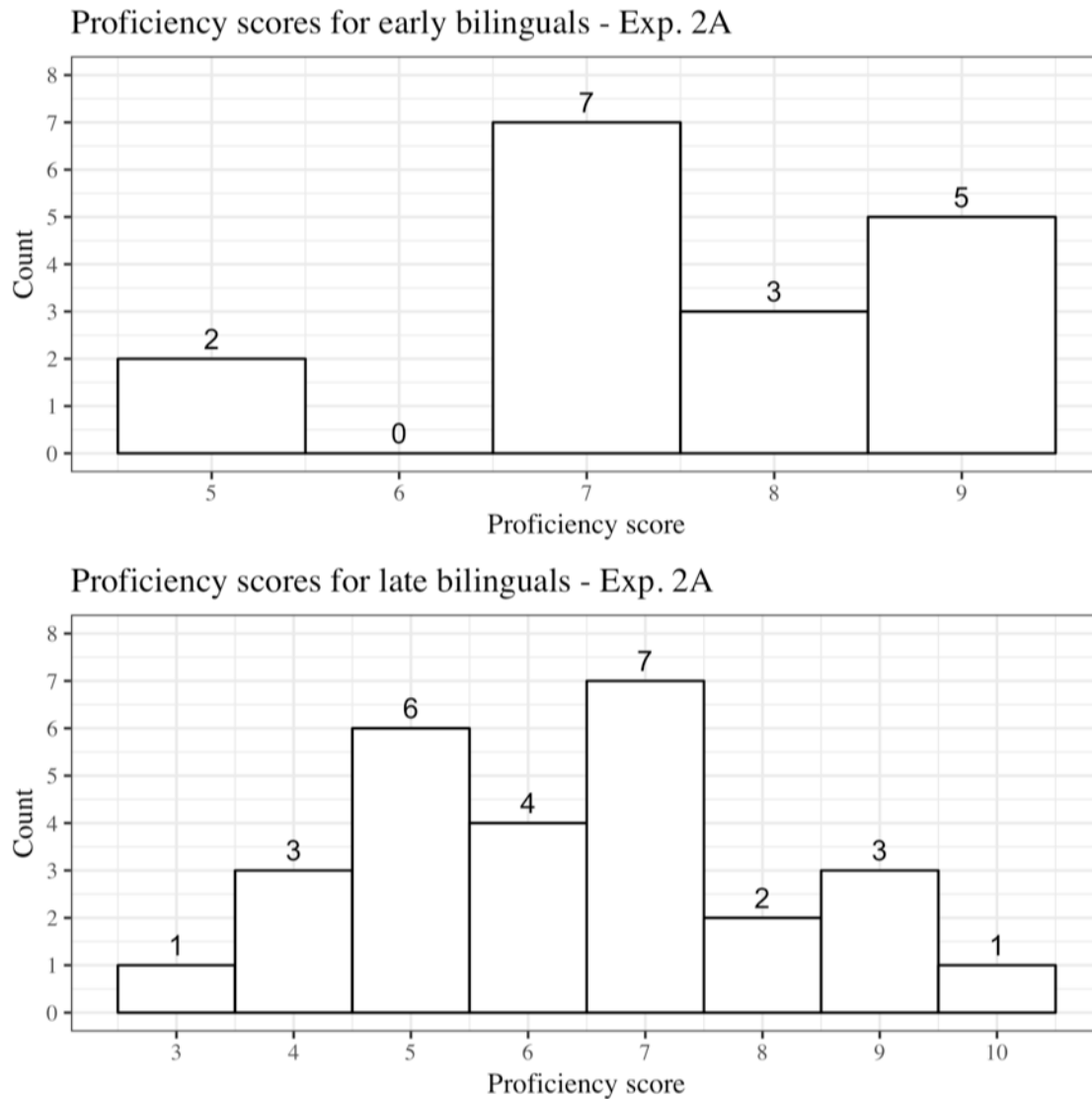


Figure 5.2. Distribution of self-reported L2 French proficiency scores out of 10 for Exp. 2a participants

The average self-reported current exposure to French as a L2 for the participants of Experiment 2a is also presented in Figure 5.3, suggesting that school and work are the two contexts in which the groups were exposed to and used French as a L2 the most.

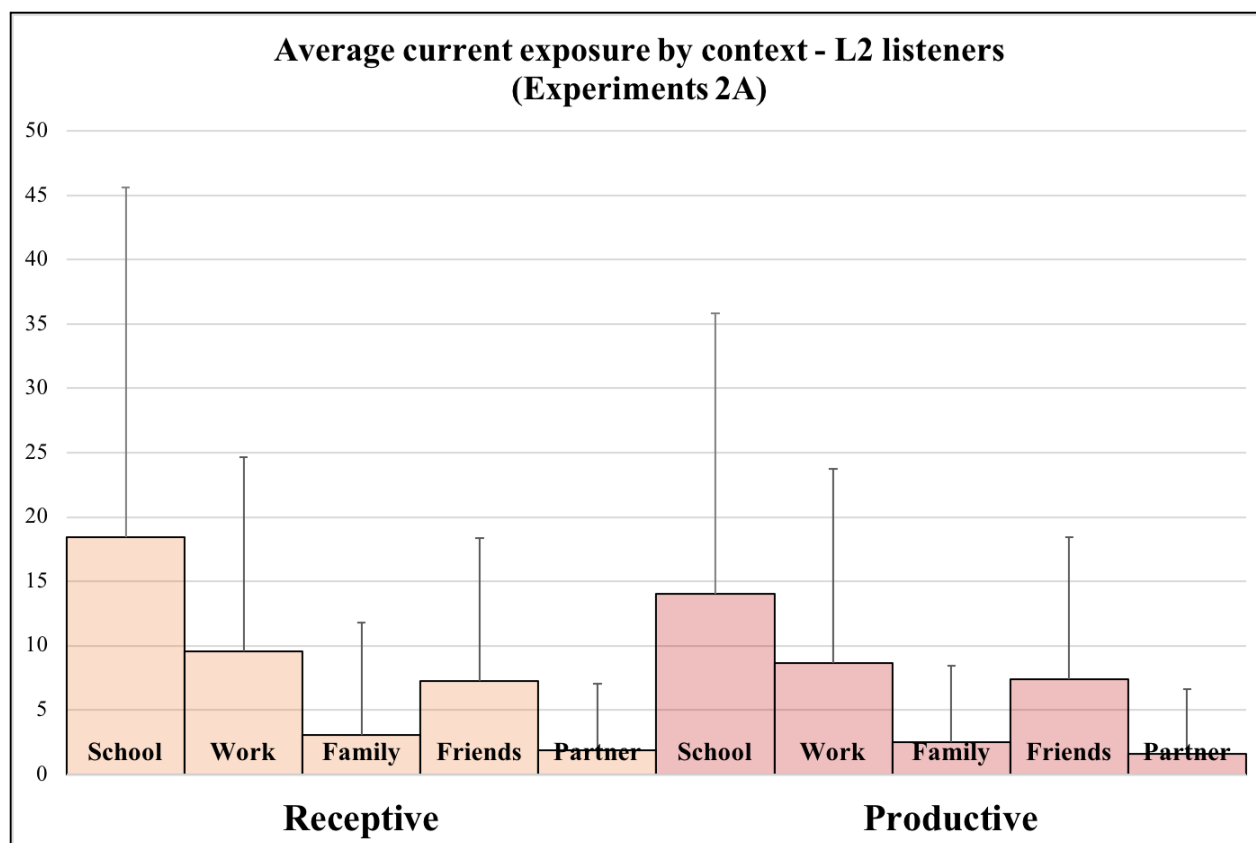


Figure 5.3. Average proportion of current exposure (receptive) to and use (productive) of L2 French in various situations for Exp. 2a participants

5.1.1.2 Stimuli and experimental conditions

The stimuli and %NASDUR values were the same as in Experiment 1a, repeated in Table 5.1. Stimuli were nine triads of monosyllabic, picturable French words containing either a contrastive nasal vowel (C $\tilde{V}$), an oral vowel in a nasalization context (i.e. followed by a nasal consonant; CVN), or an oral vowel followed by an oral consonant (CVC); see Table 5.2. Three triads contained the front-mid vowel [ɛ- $\tilde{e}$], three contained low [a- $\tilde{a}$], and three contained back-rounded [ɔ- $\tilde{o}$].

Table 5.1

Summary of the experimental conditions (%NASDUR) – Exp. 2a

		Vowel duration		
Condition	0N	Oral (100%)		50 ms nasal cons.
	20N	Oral (80%)	Nasal (20%)	
	50N	Oral (50%)	Nasal (50%)	
	80N	Oral (20%)	Nasal (80%)	
	100N	Nasal (100%)		

Note that for L2 listeners, a number of C $\tilde{V}$ -CVN words were cognates in English (e.g. *camp, vin, veine, canne, dame, rhum, tonne*). It is thus possible that these words could be recognized faster than noncognates (Sherkina, 2003), although the exact influence of the cognate status on low-level L2 phonetic processing is unclear. In the current analyses, I did not consider the words' cognate status since the number of words was limited (i.e. nine words) and the main purpose of the study was to assess the influence of fine-grained variations on recognition.

Table 5.2

Word list (experimental stimuli)

Contrastive nasal vowel	IPA	Frequency (log)	Coarticulated vowel	IPA	Frequency (log)	Fully oral vowel	IPA
<i>bain</i> ‘bathtub’	[bɛ̃]	1.703	<i>beigne</i> ‘donut’	[bɛɲ]	-0.319	<i>bec</i> ‘beak’	[bɛk]
<i>pain</i> ‘bread’	[pɛ̃]	1.798	<i>peigne</i> ‘comb’	[pɛɲ]	0.783	<i>pêche</i> ‘peach’	[pɛʃ]
<i>vin</i> ‘wine’	[vɛ̃]	1.908	<i>veine</i> ‘vein’	[vɛɲ]	1.031	<i>verre</i> ‘glass’	[vɛʁ]
<i>camp</i> ‘campsite’	[kɑ̃]	2.025	<i>canne</i> ‘cane’	[kan]	1.009	<i>cave</i> ‘basement’	[kav]
<i>dent</i> ‘tooth’	[dɑ̃]	1.123	<i>dame</i> ‘lady’	[dam]	1.937	<i>datte</i> ‘date (fruit)’	[dat]
<i>gant</i> ‘glove’	[gɑ̃]	0.994	<i>gamme</i> ‘musical scale’	[gam]	0.427	<i>gare</i> ‘bus/train station’	[gaʁ]
<i>pont</i> ‘bridge’	[pɔ̃]	1.703	<i>pomme</i> ‘apple’	[pɔm]	1.296	<i>porc</i> ‘pork’	[pɔʁ]
<i>rond</i> ‘circle’	[ʁɔ̃]	1.179	<i>rhum</i> ‘rum’	[ʁɔm]	0.799	<i>robe</i> ‘dress’	[ʁɔb]
<i>thon</i> ‘tuna’	[tɔ̃]	0.753	<i>tonne</i> ‘ton’	[tɔn]	0.651	<i>torse</i> ‘torso’	[tɔʁs]

5.1.1.3 Procedure

The experiment programming and presentation was the same as in Experiment 1a. Both eyetracking data, and RT data were collected. As in Experiment 1a, there was an error in programming meaning that the clicked image data could not be compiled properly when some participants chose the image corresponding to the oral word (CVC) or distractor.¹⁴

5.1.1.4 Analysis and variables

Chosen images, RTs and proportions of fixations to the coarticulated (CVN) image, which was considered the best target for Experiment 2a, were modeled using generalized additive mixed models (GAMMs; Wood, 2017). Details about this statistical technique can be found in Section 3.1.1.4 above (Experiment 1a), but recall that it enables one to model and assess the influence of numerical and factorial variables on potentially nonlinear, time-dependent data such as proportions of fixations.

Several variables were considered in the statistical models for the GAMM analyses in Experiment 2a. They were the same as in Experiment 1a, except for L2 FRENCH PROFICIENCY and PROVINCE OF ORIGIN. In Experiment 2a, the L2 proficiency score concerned *French* as an L2 (while it concerned English as an L2 with L1 French speakers in Experiments 1a and 1b). In addition, the PROVINCE OF ORIGIN factor was not included in the analyses of Experiment 2a, as some provinces included only one participant, therefore the sample was not representative for all provinces of origin. The assessed variables are repeated and described below, and summarized in

¹⁴ Combining participants from Experiment 2a and 2b for whom the corrected version of the experiment was used (i.e. their responses were correctly compiled, $N=68$), the CHOSEN IMAGE measure matches their given responses 96.9% of cases. This allows me to be confident that the CHOSEN IMAGE measure captures the lexical choice that was made.

Table 5.4 and Table 5.6 before the actual presentation of the results. The main variables of interest were:

- PROPORTION OF NASALIZATION DURATION ON THE VOWEL (%NASDUR): numerical value between 0 and 100, corresponding to the proportion of the vowel duration that was nasalized (0, 20, 50, 80, 100 corresponding to 0N, 20N, 50N, 80N, and 100N). This is the main manipulation of this research, varying the duration of nasalization on the vowel. Note that this variable was mean-centered by subtracting the mean from each value in order to “increase numerical stability and remove any spuriously high correlation that may arise” (Tremblay & Newman, 2015, p. 128). For example, the 0N %NasDur corresponded to -50 in the data, and the 50N %NasDur corresponded to 0 in the data.
- CHOSEN IMAGE for RT analysis: factor with two levels – Coarticulated image corresponding to the picture of the CVN word, and Nasal image corresponding to the picture of the C $\tilde{V}$ word. This was measured based on the image that was fixated the most in the last 1000 ms of the trial. This variable was ordered and treatment-coded to assess the difference between the two levels of the predictor in the GAMMs.
- VOWEL QUALITY: factor with three levels – [a], [e], [o] roughly corresponding to the quality of the vowel for each word pair. This was specifically analyzed based on previous literature on differences in nasalization between vowel qualities (Carignan, 2014; Desmeules-Trudel & Brunelle, 2018). This variable was dummy-coded as it contains more than two levels. Assessment of the differences across levels was based on the [a] baseline.

- Self-reported L2 FRENCH PROFICIENCY: numerical value out of 10, including both productive and receptive proficiency in French.
- EXPOSURE to French as a second language before school age (EARLY L2 EXPOSURE): binary factor that was coded as “1” (i.e. representing “yes”) if the participant had been exposed to English as an L2 30% of the time or more between the ages of 0 and 5 years old within the family, in kindergarten or with peers (see Note 13 in Section 5.1.1.1) and “0” (i.e. representing “no”) if they had been exposed to English less than 30% of the time in any context. This variable was ordered and treatment-coded.
- TIME: between 200 and 1300 ms after vowel onset, for the GAMM on eye tracking data only. This time window was chosen because 200 ms are necessary for eye movement planning (Fischer, 1992) after vowel onset. The end of the time window at 1300 ms was chosen based on the highest proportion of fixations value to the coarticulated image reached in the 0N condition. Time was converted into a 0-1100 interval for the statistical model, where 0 is time-locked to the vowel onset, plus 200 ms for eye movement planning.

In the presentation of the results below (Section 5.1.2), focus will be on these preceding effects (%NASDUR, CHOSEN IMAGE for RTs, VOWEL QUALITY, PROFICIENCY, EXPOSURE, TIME for eye tracking data when part of an interaction) in the main text. However, other variables were considered in the models. While these variables are interesting and can provide insights into language processing, a full discussion of all of the variables distracts from the primary goals of the research. Rather than presenting these analyses here, illustrations of the significant factors of the supplementary factors can be found in Appendix 3. Interpretation of statistical analyses of

individual listeners would not be of theoretical interest to the current work. The factors that are presented in Appendix 3 are:

- FREQUENCY OF THE NASAL (C $\tilde{V}$) WORD: numerical value, from the film subtitles subcorpus of the *Lexique* database (New et al., 2007), log-transformed and mean-centered (Tremblay & Newman, 2015).
- FREQUENCY OF THE COARTICULATED (CVN) WORD: numerical value, from the film subtitles subcorpus of the *Lexique* database (New et al., 2007), log-transformed and mean-centered (Tremblay & Newman, 2015).
- TRIAL NUMBER: numerical value between 1 and 160, which reflects participants' performance throughout the experiment, revealing anticipation effects.

5.1.2 Results

This section presents the results of Experiment 2a, focusing on the chosen images and reaction times (RTs) in Section 3.1.2.1. The eye tracking results are presented in Section 3.1.2.2.

5.1.2.1 Chosen images and reaction times

In the upper panel of Figure 5.5, participants chose the coarticulated image (CVN; between 55.33% and 75.18%) more than the nasal image (C $\tilde{V}$; between 24.82% and 44.67%) overall, similarly to L1 listeners in Experiment 1a, after rejecting the trials in which listeners reported not knowing one of the members of the pair in the familiarization (i.e. 830/3960 were rejected)¹⁵. Of the remaining 3130 trials, 2904 C $\tilde{V}$ and CVN image choices were included in the

¹⁵ Although the number of rejected trials seems high in this case, it was impossible to analyze the discrimination abilities of listeners in spoken words if one of the members was

analysis (also see Figure 5.8). The lower panel presents the GAMM predictions, and the output of the model is presented in Table 3.4. The GAMM analysis suggests that when compared to the baseline, the probability of choosing the C $\tilde{V}$ image in the 50N %NASDUR condition is significantly different (see Table 3.4) from 0N and 20N %NasDur, but not 80N and 100N. This suggests that the probability of participants choosing the C $\tilde{V}$ image increased between 0N-20N and 50N, but not when the vowel was nasalized for longer. This conforms to a categorical pattern of recognition. Additional effects are illustrated in Appendix 3.

unknown by the L2 listeners. Also, if a listener did not know one of the eighteen words, it automatically led to the rejection

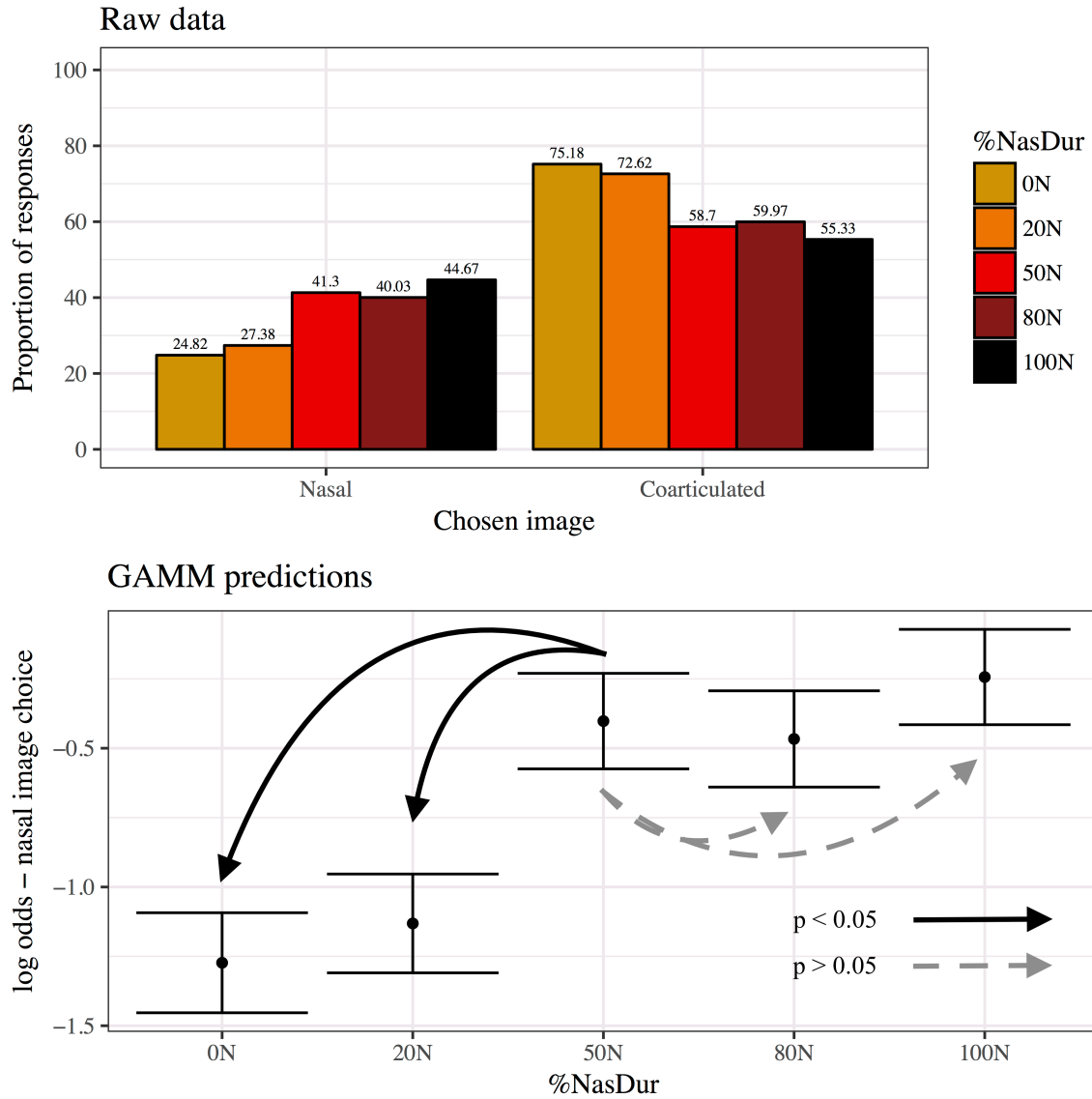


Figure 5.4. Proportions of image choices for nasal (C $\tilde{V}$) and coarticulated (CVN) images (upper panel) and GAMM predictions for image choice analysis (i.e. probability of C $\tilde{V}$ image choice by %NASDUR) – Exp. 2a

Table 5.3

GAMM results – Exp. 1a image choices

A. parametric coefficients	Estimate	Std. Error	t-value	p-value	
Intercept	-0.6064	0.2782	-2.1799	0.0293	*
Vowel_[e]	1.5910	0.1928	8.2513	< 0.0001	***
Vowel_[o]	-1.0041	0.5612	-1.7893	0.0736	
%NasDur[0N]	-0.8674	0.1389	-6.2435	< 0.0001	***
%NasDur[20N]	-0.7365	0.1361	-5.4123	< 0.0001	***
%NasDur[80N]	-0.0853	0.1295	-0.6588	0.5100	
%NasDur[100N]	0.1410	0.1281	1.1009	0.2710	
B. smooth terms	edf	Ref.df	F-value	p-value	
s(Freq_coart.)	4.3697	4.7436	113.3608	< 0.0001	***
s(Freq_nasal)	1.0000	1.0001	8.0135	0.0046	**
Random_Participant	34.5941	43.0000	163.6797	< 0.0001	***
Random_Item	0.0023	4.0000	0.0006	0.4091	

The left panel in Figure 5.5 shows that when participants chose the nasal (C $\tilde{V}$) image, there was an inverse relationship between RT and the duration of nasalization on the vowel. However, when participants chose the coarticulated (CVN) image (right panel), RT was the longest when the vowel was not nasalized at all. For the other %NASDUR values, RTs seem similar for CVN (i.e. they hover around 1900 ms on average), perhaps with a small increase in RT values when vowels were nasalized between 50% and 100% of their duration.

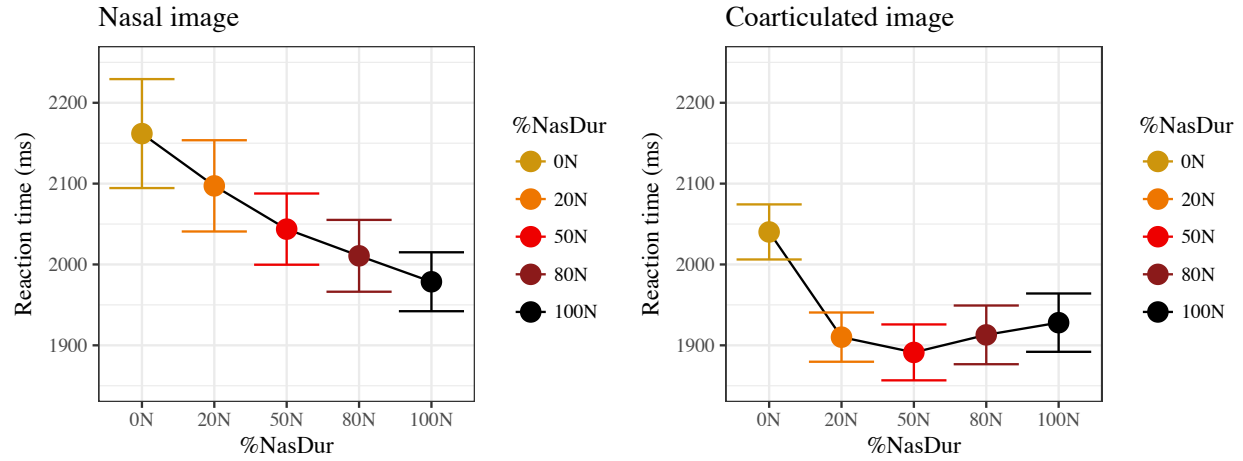


Figure 5.5. RT averages for selected nasal (C $\tilde{V}$) images by %NASDUR (left) and coarticulated (CVN) images (right) by %NASDUR – Exp. 2a

The remainder of this section presents the results of the GAMM analysis of RTs (for a similar model fitting procedure, see Porretta et al., 2016). The input variable of the model was the log-transformed RT. Random effects structure was determined by maximum-likelihood chi-square tests (*compareML()* function from the *itsadug* package), therefore random smooth (i.e. nonlinear) terms by PARTICIPANTS and %NASDUR, and by ITEM (minimal pair label) and %NASDUR were kept in. Predictors of interest and interactions that were assessed are summarized in Table 5.4.

Table 5.4

Summary of the independent variables for the full RT GAMM – Exp. 2a

Parametric terms	Levels
CHOSEN IMAGE	Factor w/ 2 levels: Nasal, Coarticulated image
VOWEL QUALITY	Factor w/ 3 levels: [a], [e], [o]
EXPOSURE TO L2 BEFORE 5	Factor w/ 2 levels: True or False
Smooth terms	Levels
%NASDUR	Numerical (centered): -50 to 50
TRIAL NUMBER	Numerical: 1 to 160
LOG FREQUENCY (CVN)	Numerical: -1.0804 to 1.1756
LOG FREQUENCY (C $\tilde{V}$)	Numerical: -0.5229 to 0.7491
L2 FRENCH PROFICIENCY	Numerical: 3 to 10
Interactions	
%NASDUR*VOWEL QUALITY	
%NASDUR*FIXATED IA	
Random effects	
Smooth term:	
PARTICIPANTS*%NASDUR	
Smooth term:	
MINIMAL PAIR LABEL*%NASDUR	

For final model building, the same backwards stepwise procedure as in the other experiments in Chapter 3 was used. Specific details about the procedure are found in Section 3.1.1.4. Results of the final model are presented in Table 5.5 with CHOSEN IMAGE and EXPOSURE TO L2 BEFORE AGE 5 as parametric factors, and significant smooth terms for %NASDUR and TRIAL. Note that p -values lower than 0.05 here suggest that the smooth term is significantly different from zero, suggesting statistical difference between two or more points in the continuum (e.g. %NASDUR or TRIAL NUMBER). Presentation of the results will focus on the significant %NASDUR smooth term, which is crucial to the research question. Illustration of the additional significant factor (TRIAL) is found in Appendix 3.

Table 5.5

GAMM results – Exp. 2a RTs

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
Intercept	3.2476	0.0150	216.4936	< 0.0001
Chosen_image	0.0144	0.0053	2.7069	0.0068
L2_earlyExp	0.0499	0.0215	2.3243	0.0202
B. smooth terms	edf	Ref.df	F-value	p-value
s(%NasDur)	1.8775	2.1136	3.6373	0.0223
s(Trial)	6.3289	7.4326	9.4227	< 0.0001
Random_Participant	44.3392	173.0000	4.7854	< 0.0001
Random_Item	19.0671	34.0000	2.6402	< 0.0001

Parametric coefficient estimates in Table 5.5 suggest that overall, when participants fixated on the nasal (C $\tilde{V}$) image, RTs were significantly different than when participants fixated the coarticulated (CVN) image (3.248 ± 0.0144 , $t = 2.707$, $p < 0.05$). This seems to match the raw RT values presented in Figure 5.5. Furthermore, RTs were different when listeners had been exposed to French before five years old versus participants who had not (3.248 ± 0.0499 , $t = 2.324$, $p < 0.05$). Model predictions (Figure 5.6) suggest that listeners in the early-exposure group were overall slower than listeners who were exposed to L2 French later in life. This is surprising, as it is expected that listeners who have been exposed more and in a naturalistic environment to L2 French would display better L2 processing abilities, therefore smaller RTs. However, better knowledge and a more integrated lexicon (linked to early and naturalistic L2 exposure; Sabourin et al., 2014) of both C $\tilde{V}$ and CVN words might induce more competition among lexical items (Dahan et al., 2001a) and cause slower RTs for listeners who were in the early-exposure group.

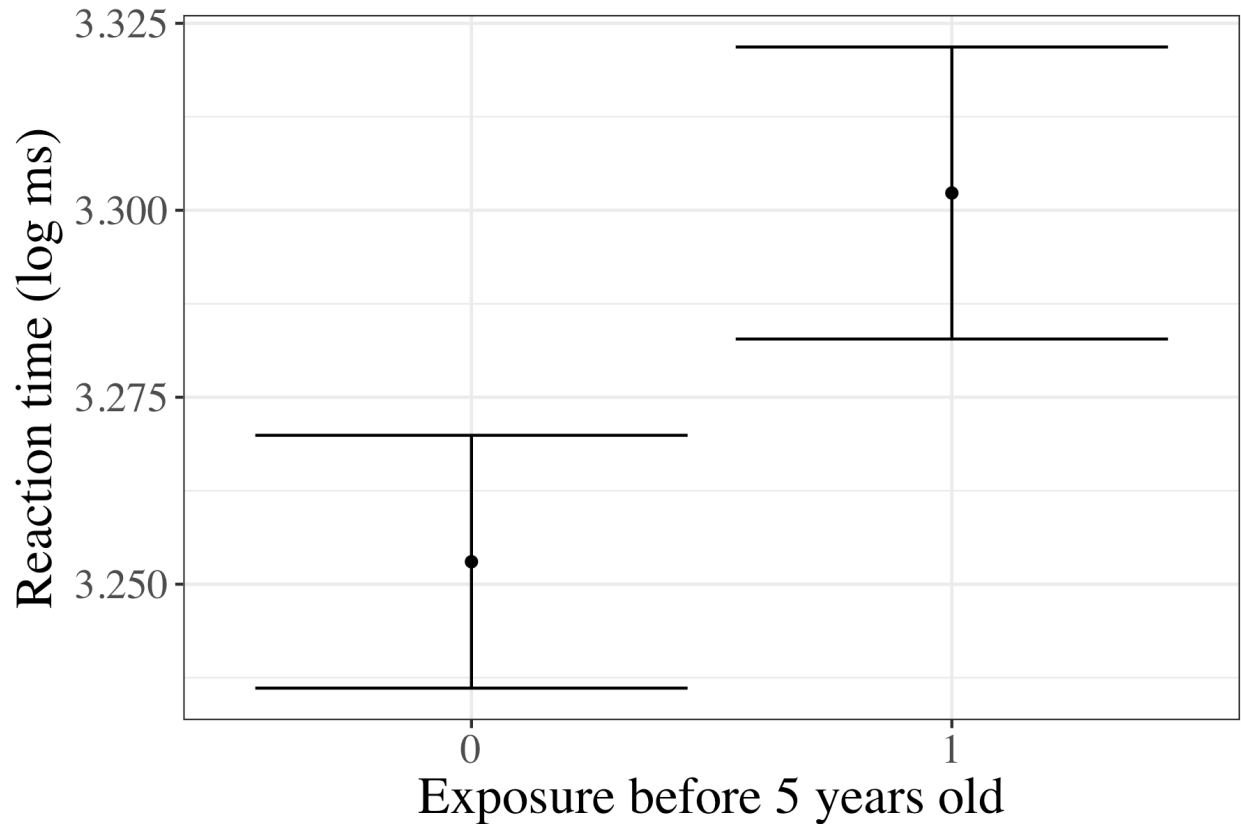


Figure 5.6. Predicted data by EARLY-EXPOSURE-GROUP – Exp. 2a RTs

Figure 5.7 presents the data that is predicted by the model according to the %NASDUR smooth term. In this Figure, RTs become shorter as the duration of nasalization on the vowel increases overall. This suggests that processing load decreases as vowel nasalization becomes longer, therefore the word that corresponds to the heard stimuli becomes more obvious for listeners. This approximates the raw RT values presented in Figure 5.5, but recall that the predicted data presented here does not take the CHOSEN IMAGE into consideration.

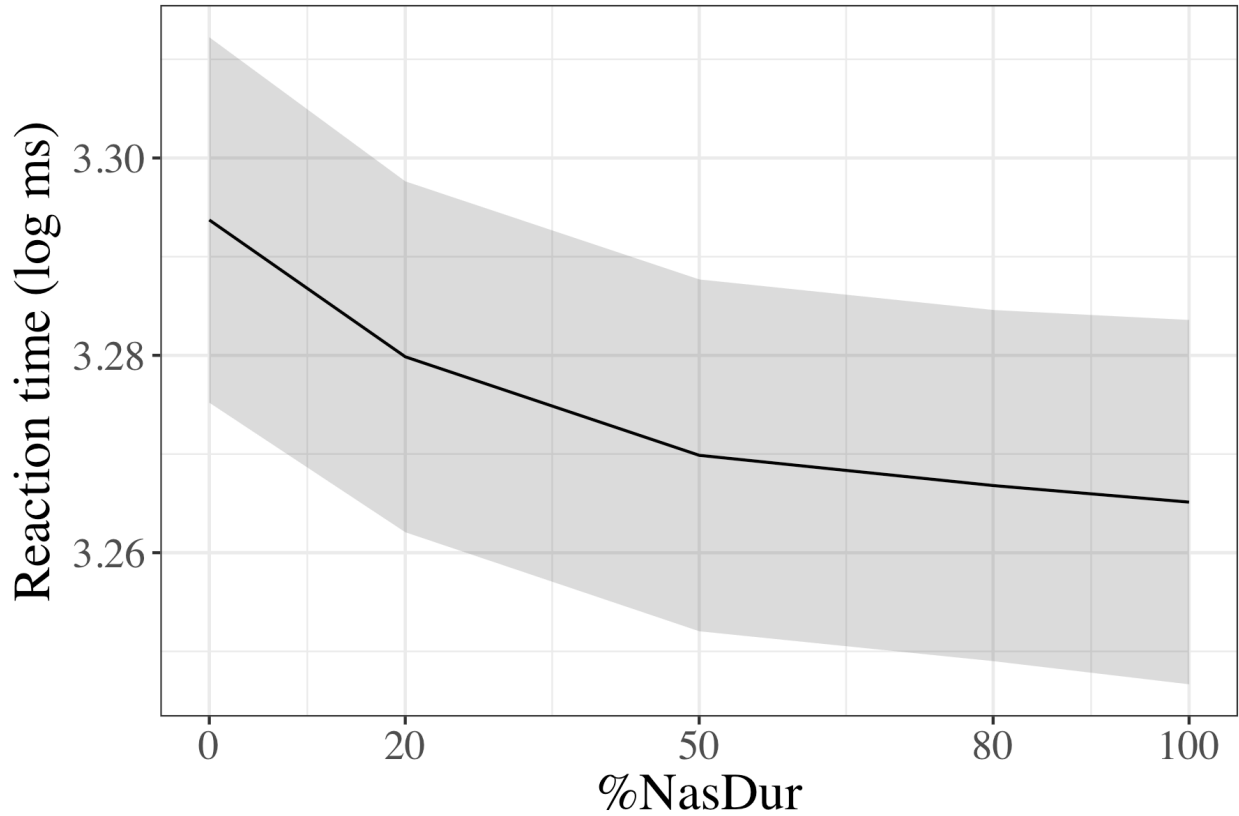


Figure 5.7. Predicted data by %NASDUR – Exp. 2a RTs

5.1.2.2 Eye movements

This subsection presents the eye tracking results, focusing on the proportions of fixations between 200 and 1300 ms after vowel onset. Similarly to Experiments 1a and 1b, I present the patterns of fixations to the C $\tilde{V}$ (leftmost panels) and CVN (rightmost panels) when listeners chose the C $\tilde{V}$ image (upper; 1038 trials) and the CVN image (lower; 1866 trials). I will start by presenting the general eye movement patterns, followed by separate GAMM analyses of C $\tilde{V}$ and CVN image choices before the discussion in Section 5.1.3.

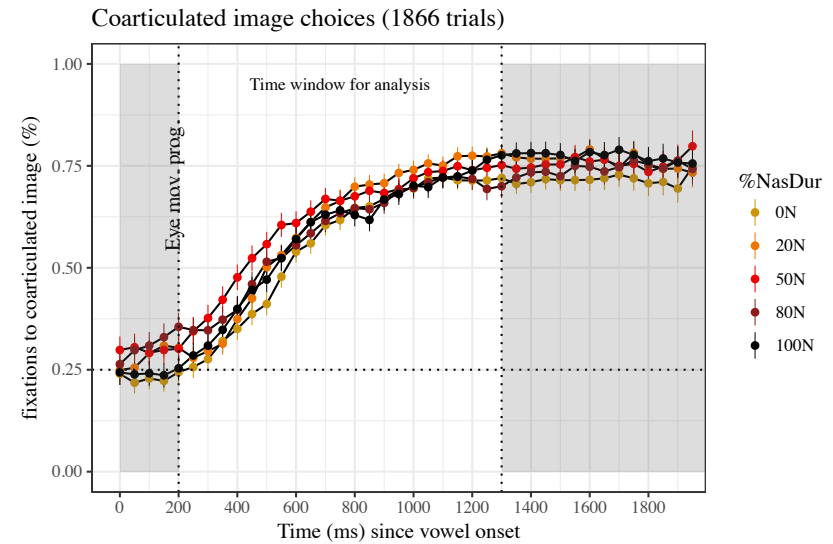
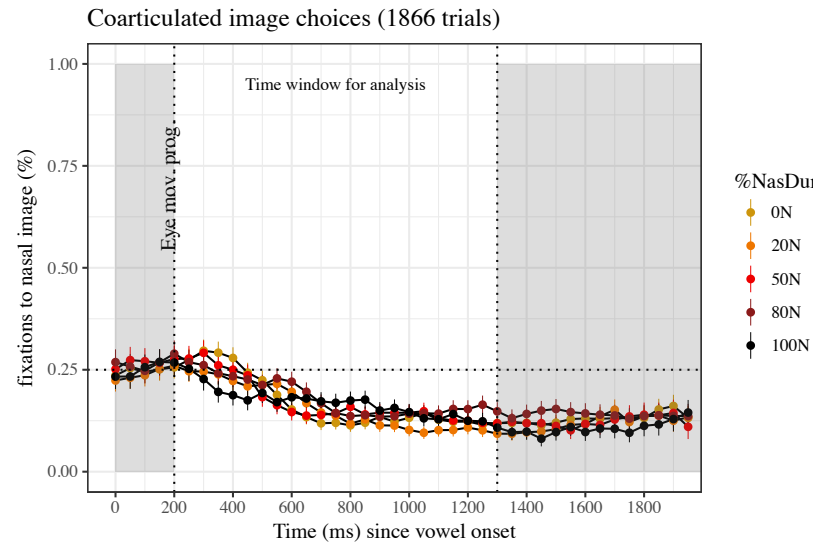
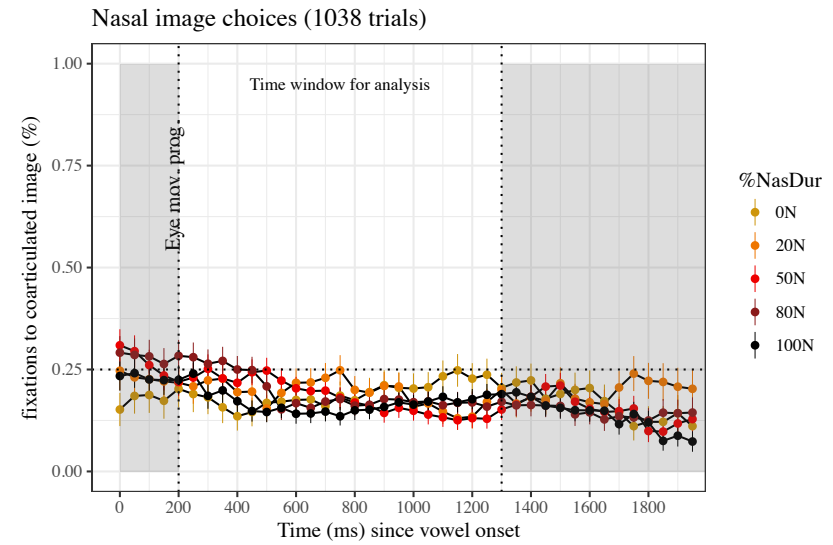
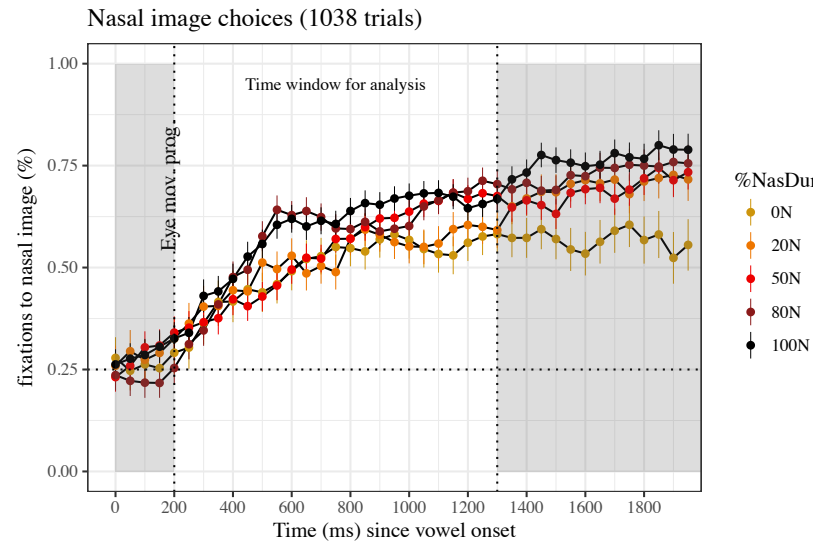


Figure 5.8. Proportions of fixations to the nasal ($C\tilde{V}$) image (leftmost) and coarticulated (CVN) image (rightmost), nasal ($C\tilde{V}$) choices (upper) and coarticulated (CVN) choices (lower) – Exp. 2a

The upper left panel of Figure 5.8 **Erreur ! Nous n'avons pas trouvé la source du renvoi.** illustrates the relationship between the proportions of fixations and the amount of nasalization on the vowel, especially at the end of the time window. Here, when vowels were nasalized for a longer period (50N or more %NASDUR), participants tended to look slightly more at the nasal (C $\tilde{V}$) item than in the other %NASDUR conditions, especially at the end of the analysis window. This suggests an influence of nasalization duration on word recognition of spoken words in L2 French. On the other hand, when participants chose the coarticulated (CVN) image, no specific pattern seems to emerge on average – all %NASDUR conditions seem relatively undifferentiated throughout their time course, except 50N at the beginning of the analysis window.

Similarly to the other experiments, the GAMM was fitted to the eye tracking data using the maximum likelihood (ML) method (see Porretta et al., 2016). The input variable was the empirical logit-transformed fixations (Barr, 2008) to the nasal (C $\tilde{V}$) and coarticulated (CVN) image, depending on the image choice. The best random effects structure included random intercepts by participant and minimal pair label. The predictors that were assessed are summarized in Table 5.6. Note that they are the same as in Experiment 1a, except for the PROVINCE OF ORIGIN factor which was included in Experiment 1a but not considered in Experiment 2a, and L2 PROFICIENCY which is the self-reported score in L2 French in the current experiment.

Table 5.6

Summary of the independent variable for the full GAMM – Exp. 2a %fix

Parametric terms	Levels
CHOSEN IMAGE	Factor w/ 2 levels: Nasal (baseline), Coarticulated image
VOWEL QUALITY	Factor w/ 3 levels: [a], [e], [o]
EXPOSURE TO L2 BEFORE 5	Factor w/ 2 levels: True or False
Smooth terms	Levels
TIME	Numerical: 0 to 1100, separated in 50 ms time bins
%NASDUR	Numerical (centered): -50 to 50
TRIAL NUMBER	Numerical: 1 to 160
LOG FREQUENCY (CVN)	Numerical: -1.0804 to 1.1756
LOG FREQUENCY (C $\tilde{V}$)	Numerical: -0.5229 to 0.7491
L2 FRENCH PROFICIENCY	Numerical: 3 to 10
Interactions	
TIME* FIXATED IA	
TIME* VOWEL QUALITY	
TIME* PROVINCE OF ORIGIN	
TIME* EXPOSURE TO L2 BEFORE 5	
TIME*%NASDUR	
TIME* TRIAL NUMBER	
TIME* LOG FREQUENCY (COART.)	
TIME* LOG FREQUENCY (NASAL)	
TIME* L2 FRENCH PROFICIENCY	
TIME*%NASDUR *VOWEL QUALITY	
TIME*CONDITION*FIXATED IA	
Random effects	
Smooth term: PARTICIPANTS*%NASDUR	
Smooth term:	
MINIMAL PAIR LABEL *%NASDUR	

The output of the final model on (empirical-logit-transformed) fixations to the nasal (C $\tilde{V}$) image when listener chose the nasal (C $\tilde{V}$) image (1038 trials) is presented in Table 5.7 with EARLY EXPOSURE as a parametric factor, significant smooth terms for TIME, TRIAL and %NASDUR, as well as a 2-way interaction between TIME and TRIAL, and a 3-way interaction between TIME, CONDITION, and VOWEL QUALITY. Results presentation will focus on the significant EARLY EXPOSURE, %NASDUR, and TIMEX%NASDURXVOWEL QUALITY factors. Illustration of the additional significant smooths (TRIAL NUMBER, and interaction between TIME and TRIAL) is found

in Appendix 3. Auto-correlation AR1 value, corresponding to the average correlation between one data point and the preceding one in the time series, was empirically determined to 0.836 based on the data. Also note that the non-significant VOWEL QUALITY parametric factor as well as the non-significant TIMEX%NASDUR interaction were kept in the model as baseline for the higher-order interaction.

Table 5.7

GAMM results – Exp. 2a %fix to C $\tilde{V}$, C $\tilde{V}$ choices

A. parametric coefficients	Estimate	Std. Error	t-value	p-value	
Intercept	0.2212	0.3516	0.6290	0.5293	
Vowel_[e]	0.3584	0.4163	0.8610	0.3892	
Vowel_[o]	-0.1883	0.4218	-0.4465	0.6552	
L2_earlyExp	-0.4614	0.2294	-2.0117	0.0443	*
B. smooth terms	edf	Ref.df	F-value	p-value	
s(Time)	3.4595	3.9407	11.4596	< 0.0001	***
s(Trial)	4.6466	5.6923	2.9583	0.0133	*
s(%NasDur)	1.0160	1.0312	31.1043	< 0.0001	***
s(Time:Trial)	9.1607	11.6457	2.4053	0.0047	**
s(Time:%NasDur)	2.1628	2.4424	2.8135	0.0805	
s(Time:%NasDur):Vowel_[a]	8.2470	11.1271	3.4173	0.0001	***
s(Time:%NasDur):Vowel_[e]	8.2895	11.1133	1.2491	0.2549	
s(Time:%NasDur):Vowel_[o]	8.5126	10.9145	2.2751	0.0093	**
Random_Participant	234.8700	394.0000	1.9617	< 0.0001	***
Random_Item	36.2696	78.0000	1.1972	< 0.0001	***

Figure 5.9 shows the data that is predicted by the GAMM regarding the general effect of L2 EXPOSURE. Overall, participants who were early-exposed to French as a L2 tended to fixate less to the nasal (C $\tilde{V}$) image. This is surprising, since it would be expected that listeners who learned French as a L2 at an earlier age should have more a robust nasal vowel ($\tilde{V}$) category, and therefore would fixate the C $\tilde{V}$ item more. However, it can also be that a more robust category

makes them overall more conservative regarding the $\tilde{V}$ category, and that only good instances of $C\tilde{V}$ stimuli will be recognized as such.

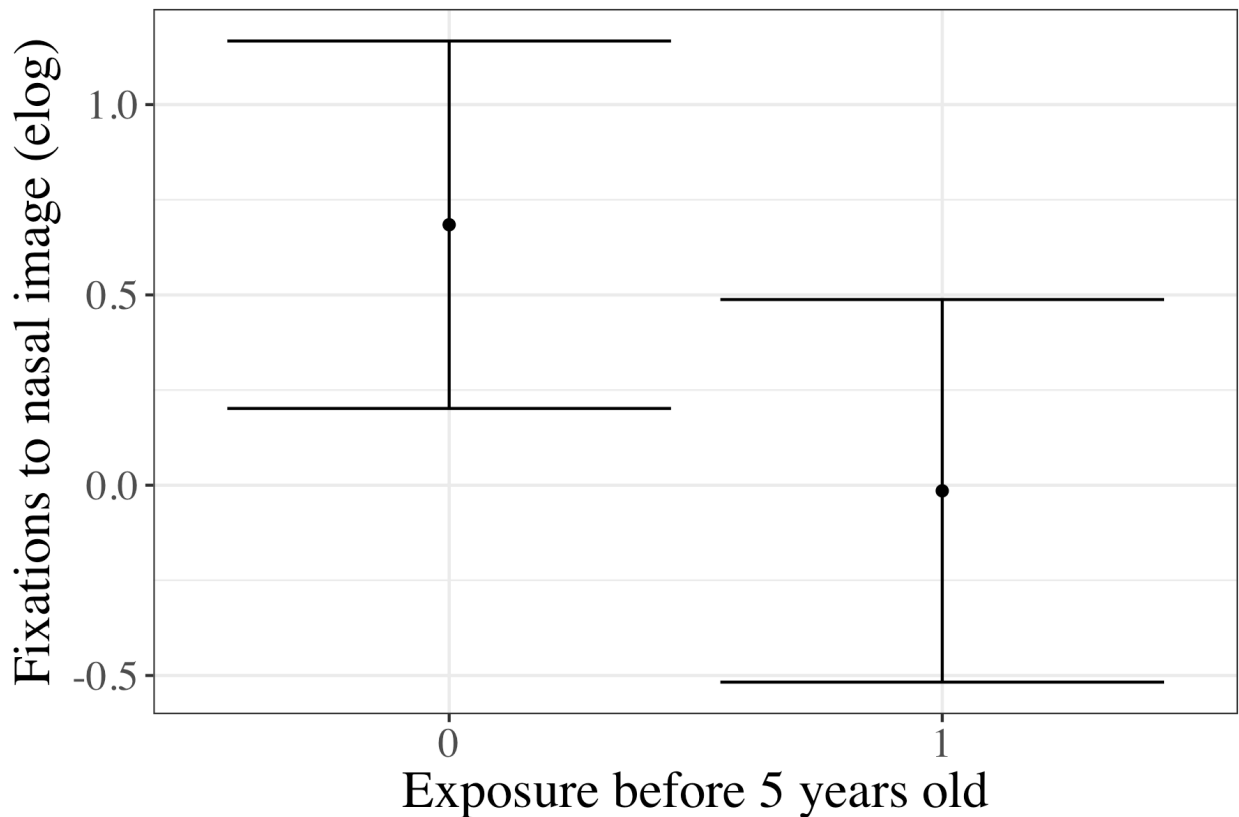


Figure 5.9. Predicted data by EXPOSURE group – Exp. 2a elog %fix to $C\tilde{V}$, $C\tilde{V}$ choices

The predicted data in Figure 5.10 shows that there are differences among %NASDUR conditions in fixations to the $C\tilde{V}$ image for $C\tilde{V}$ choices. In this case, smaller %NASDUR values yield fewer fixations to the nasal image, and inversely for higher %NASDUR values. On the surface, the increase seems gradient, similarly to L1 listeners in Experiment 1a. However, looking at the raw data in the upper left panel of Figure 5.8, it seems like the 50N %NASDUR value behaves with the 0N-20N %NASDUR conditions at the beginning of the time window, and with 80N-100N for the second half of the time window. Thus, it is likely that the 50N %NASDUR

value is at the source of the apparent gradient pattern. It is thus difficult to draw a firm conclusion on the pattern type in this case, and recognition can be considered more categorical with L2 speakers with ambiguous stimuli.

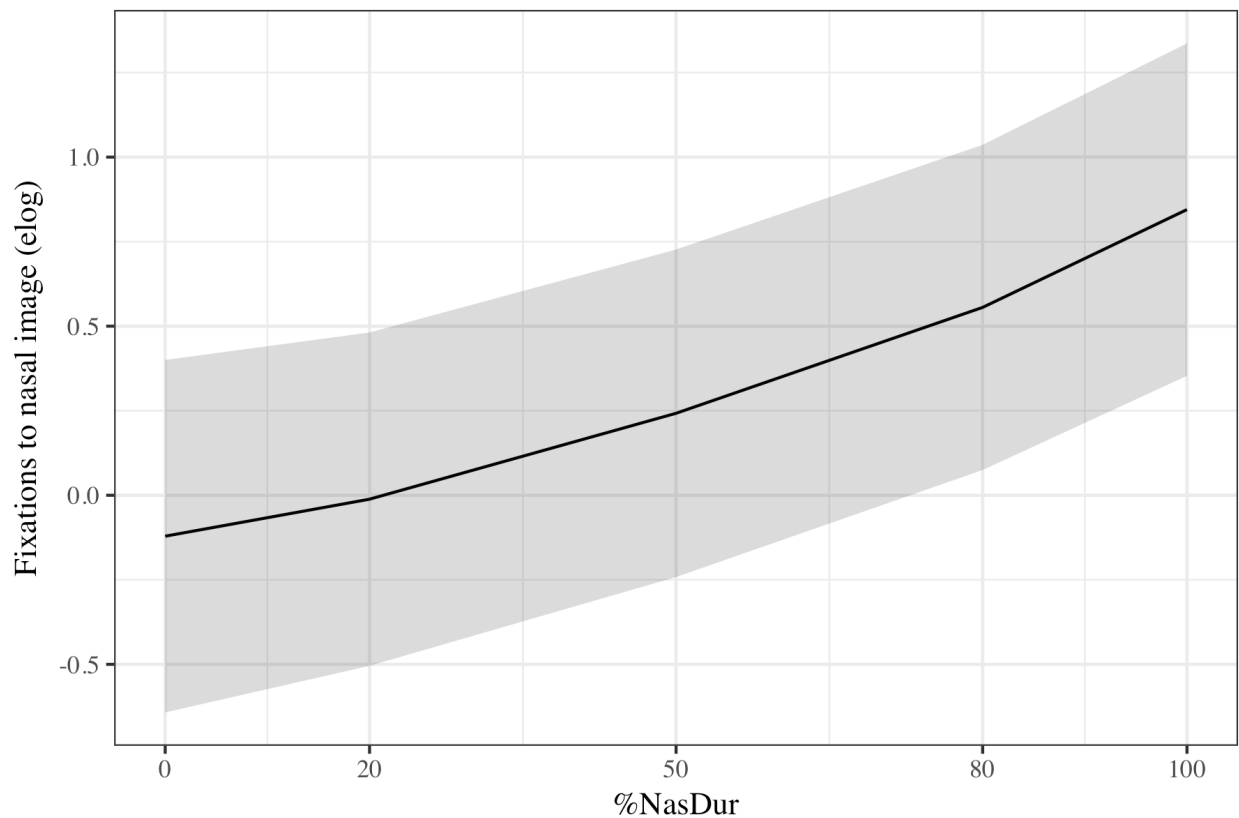


Figure 5.10. Predicted data %NASDUR – Exp. 2a elog %fix to C $\tilde{V}$, C $\tilde{V}$ choices

Figure 5.11 shows a complex pattern of fixations to C $\tilde{V}$ through TIME depending on both the %NASDUR condition and QUALITY OF THE VOWEL. For example, the fixation pattern for the 0N %NASDUR value is different through time for all three vowels – increase and decrease for [a] stimuli, gradual increase for [e] stimuli, and overall decrease for [o] stimuli. Similarly to the other experiments presented in Chapter 3, this significant interaction is a testament of the complexity

of processing of CF vowel nasalization, but it remains difficult to draw firm conclusions regarding vowel quality and duration of nasalization.

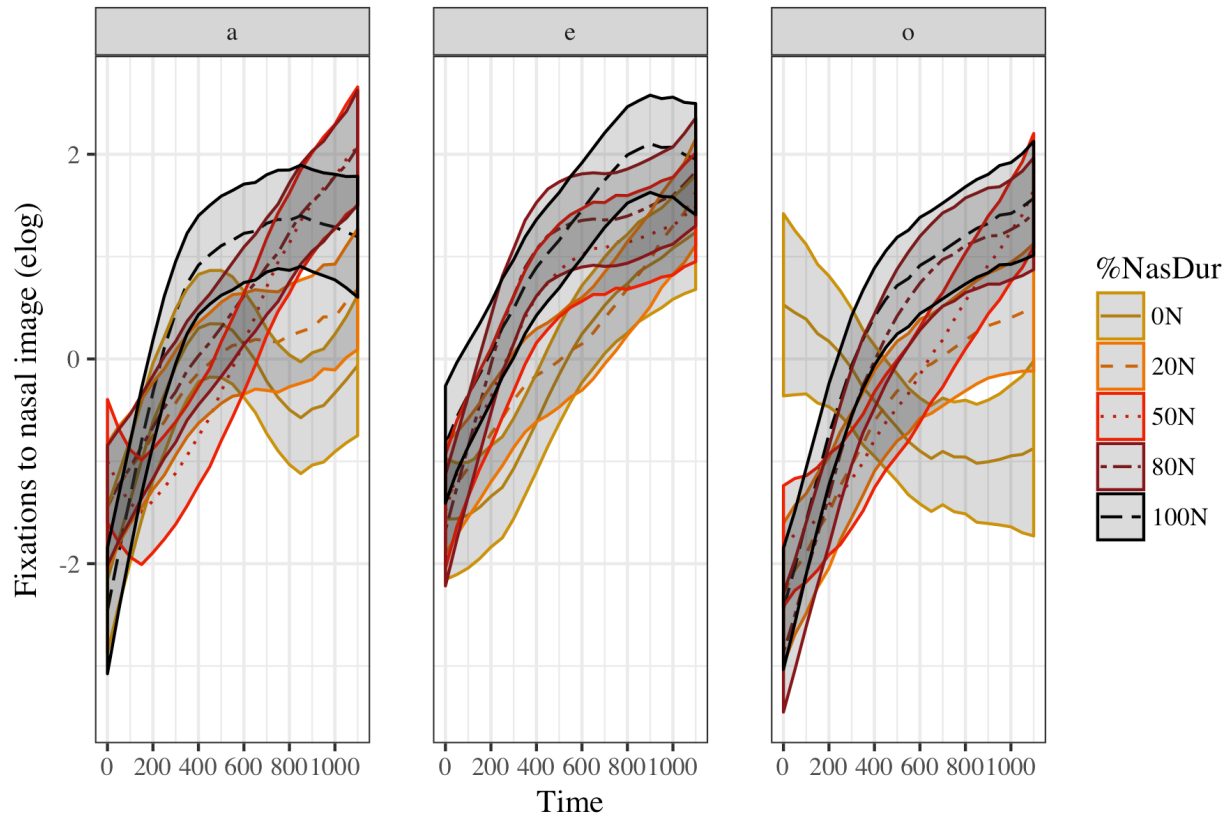


Figure 5.11. Predicted data by %NASDUR and VOWEL QUALITY through TIME – Exp. 2a elog %fix to C $\tilde{V}$, C $\tilde{V}$ choices

The second GAMM modeled (empirical-logit-transformed) proportions of fixations to the coarticulated (CVN) image when participants chose the CVN image (1866 trials). The model output is presented in Table 5.8. Presentation of the results will focus on the EXPOSURE factor, the %NASDUR smooth ($p = 0.053$) as well as the TIME $\times$ VOWEL QUALITY and TIME $\times$ %NASDUR $\times$ VOWEL QUALITY interactions. The remaining significant TIME, TRIAL NUMBER,

FREQUENCY (COARTICULATED) factor, and TIMEXTRIAL NUMBER and TIMEXFREQUENCY (NASAL) interactions are illustrated in Appendix 3.

Table 5.8

GAMM results – Exp. 2a %fix to CVN, CVN choices

A. parametric coefficients	Estimate	Std. Error	t-value	p-value	
Intercept	0.4793	0.2493	1.9228	0.0545	
Vowel_[e]	0.0395	0.3193	0.1239	0.9014	
Vowel_[o]	0.6676	0.2874	2.3230	0.0202	*
L2_earlyExp	-0.4212	0.1678	-2.5105	0.0121	*
B. smooth terms	edf	Ref.df	F-value	p-value	
s(Time)	4.4790	4.9988	4.8628	0.0002	***
s(Trial)	3.3468	4.1449	8.2504	< 0.0001	***
s(%NasDur)	2.3984	2.7397	2.2207	0.0532	
s(Freq_coart)	1.6857	1.7916	4.6163	0.0385	*
s(Freq_nasal)	1.0070	1.0093	0.2243	0.6353	
s(Time):Vowel_[a]	1.0109	1.0140	0.5929	0.4398	
s(Time):Vowel_[e]	0.0276	0.0345	0.0945	0.9544	
s(Time):Vowel_[o]	1.0057	1.0072	4.0582	0.0434	*
s(Time: Trial)	9.9857	12.7406	2.3765	0.0041	**
s(Time:%NasDur)	1.0130	1.0240	0.7966	0.3738	
s(Time:Freq_nasal)	2.0900	2.3582	3.2905	0.0325	*
s(Time:%NasDur):Vowel_[a]	0.9608	1.4370	1.4009	0.1871	
s(Time:%NasDur):Vowel_[e]	2.9758	3.5972	3.1754	0.0219	*
s(Time:%NasDur):Vowel_[o]	10.1717	13.1386	3.2149	0.0001	***
Random_Participant	229.6146	394.0000	2.1520	< 0.0001	***
Random_Item	48.9588	76.0000	2.7244	< 0.0001	***

The significant L2 EARLY EXPOSURE factor in Table 5.8 is illustrated in Figure 5.12, and suggests differences between L2 listeners who were significantly exposed to French before five years-old (i.e. “1” level) and those who were not (i.e. “0” level). The predicted data suggests that listeners who are in the early exposure group tended to fixate slightly less the coarticulated (CVN) image. Early within the time window, early-exposed listeners tended to fixate the CVN image less. This reflects that they may more easily recognize words that were nasalized for a

longer duration as C $\tilde{V}$, which would suggest more robust representations of phonological nasal (C $\tilde{V}$) vowels. This is more or less consistent with the exposure effect on fixations to the C $\tilde{V}$ image above (Figure 5.9), since early bilinguals fixated less both the coarticulated (CVN) and nasal (C $\tilde{V}$) images overall. However, if one assumes that their nasal VS. coarticulated categories are more robust overall, it may be the case that they tend to be more conservative in general in making decisions about words in L2 French.

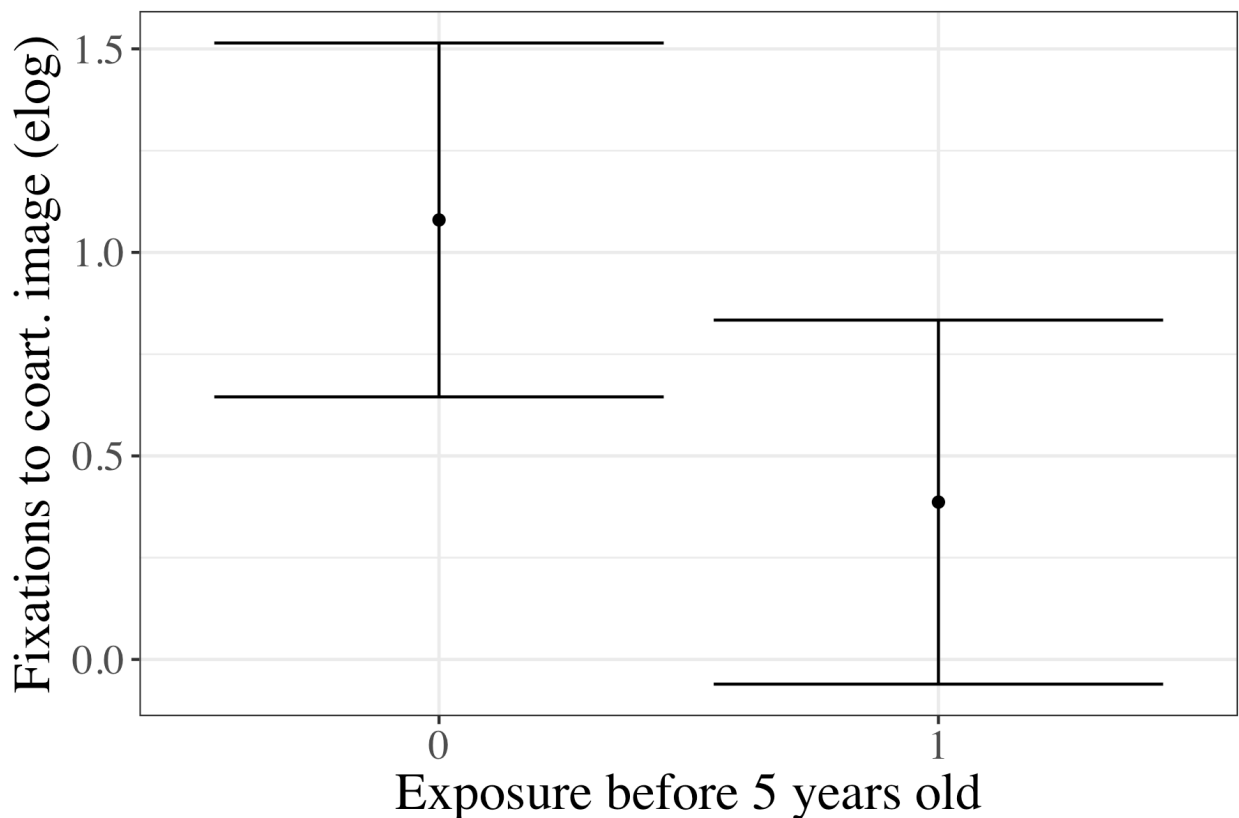


Figure 5.12. Predicted data by EXPOSURE – Exp. 2a elog %fix to CVN, CVN choices

The data predicted by the (almost) significant %NASDUR smooth, which had a p -value of 0.053 in the final GAMM, is presented in Figure 5.13. In the figure, 0N %NASDUR yielded lower proportions of fixations to the coarticulated image than the 20N and 50N %NASDUR. For higher

%NASDUR, proportions of fixations to the CVN image also seem lower than 20N and 50N.

However, looking at the y-axis, the average values are all between 0.6 and 1 on the empirical logit scale, thus suggesting small overall variations cross %NASDUR values. Consequently, even though the patterns of fixations seem to conform to a categorical pattern of recognition, the high p -value and small range of average values cannot be interpreted as firm evidence in favour of categorical or gradient recognition in this case.

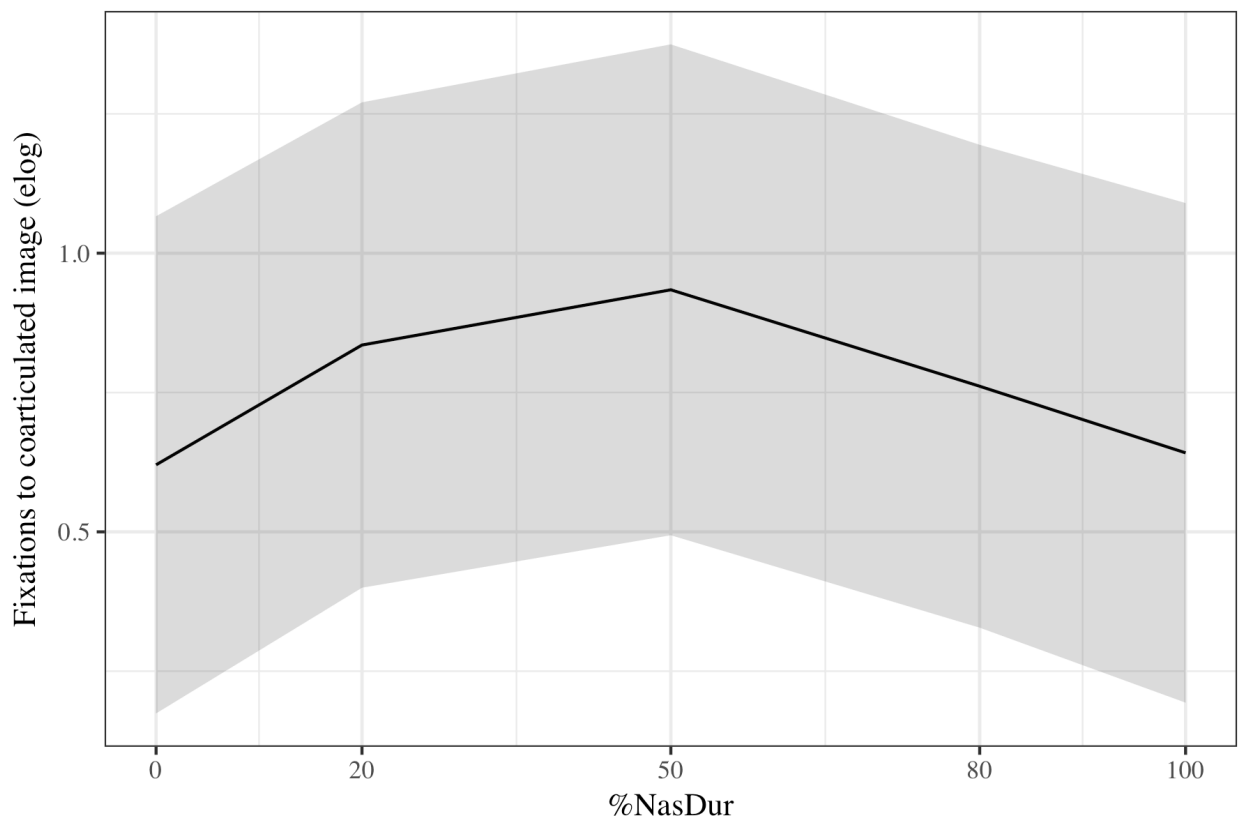


Figure 5.13. Predicted data by %NASDUR– Exp. 2a elog %fix to CVN, CVN choices

The predicted data by the significant TIMEXVOWEL QUALITY interaction is presented in Figure 5.14. It is shown that overall through time, [o] stimuli yielded more fixations to the coarticulated (CVN) image than the [e] stimuli, and [a] stimuli yielded intermediate patterns of

fixations compared to the other two vowel qualities. This can be due to the *pomme* stimulus, which was the best-known word in the familiarization phase overall by the group of listeners. If listeners saw a very well-known image (e.g. *pomme*) on the display, it is likely that they will tend to fixate it more than when they are less familiar with the image.

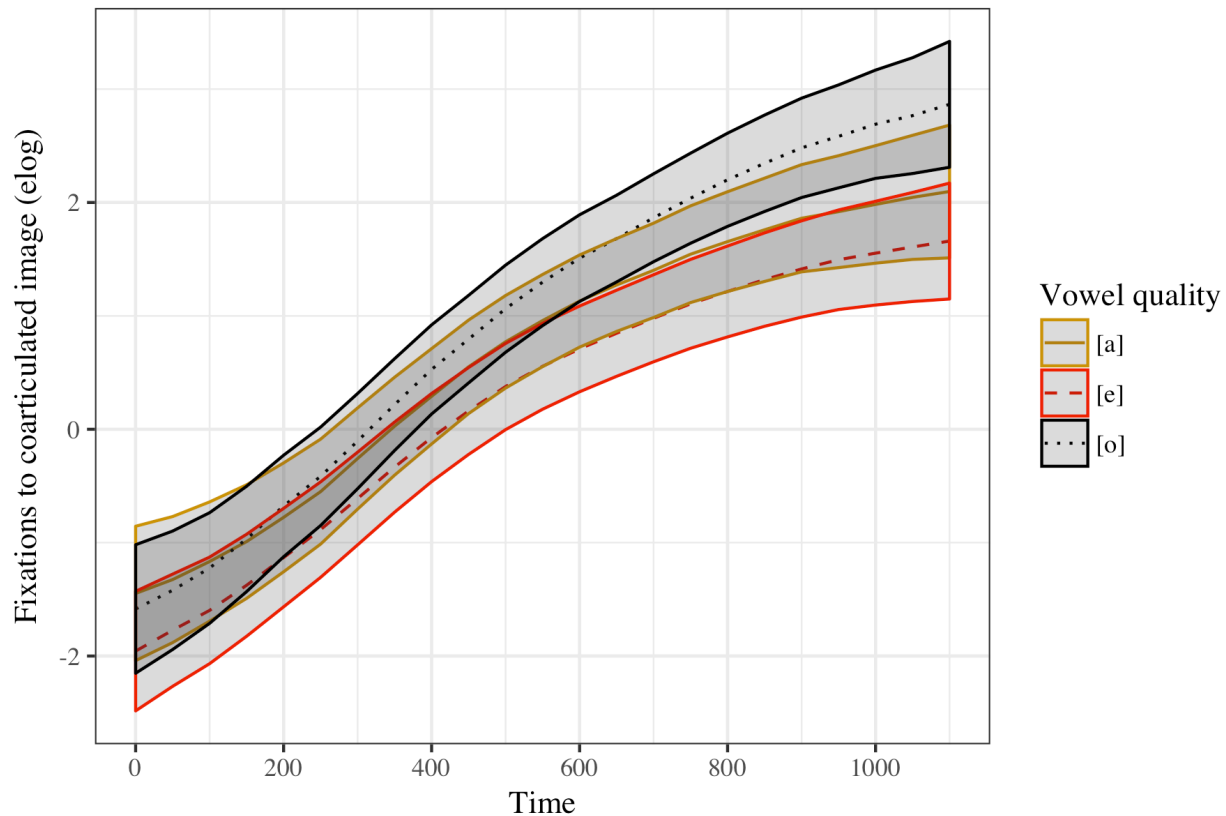


Figure 5.14. Predicted data by VOWEL QUALITY through TIME – Exp. 2a elog %fix to CVN, CVN choices

Finally, Figure 5.15 shows the predicted data for the 3-way interaction between TIME, %NASDUR and VOWEL QUALITY. The data suggests that for [ɔ-õ], differences across %NASDUR values are slightly greater through TIME than for [a-ã] and [ɛ-ẽ]. These differences within [ɔ-õ] words were also found for L1 speakers (Figure 3.11). Although the reasons why [ɔ-õ] words

behave so differently are not clear and the pattern complexity was present for all other analyses presented above, it is likely that individual words and the participants' familiarity played a role in the emergence of this pattern. This fits within the SLM and PAM-L2 models of L2 perceptual learning (Best & Tyler, 2007; Flege, 1995), which suggest that exposure to specific sounds will increase the robustness of the L2 category, and therefore increase discrimination performance on the contrast.

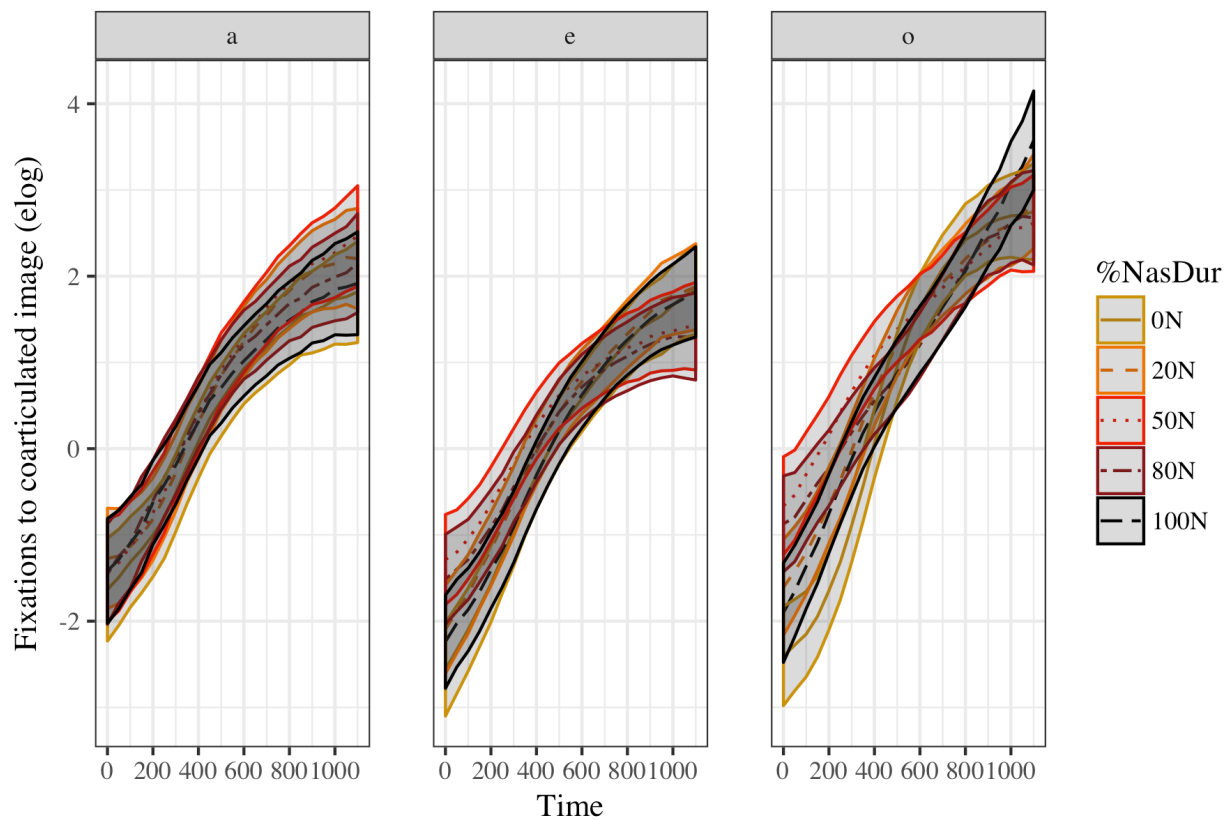


Figure 5.15. Predicted data by VOWEL QUALITY and %NASDUR through TIME – Exp. 2a elog %fix to CVN, CVN choices

In any psycholinguistic experiment using real words, it is expected that some differences will emerge across individual items. Furthermore, it is reasonable to say that these cross-item

differences may be exaggerated when investigating L2 listeners, especially since it is difficult to control for external variables such as L2 proficiency or experience, and actual knowledge of individual words in less-experienced listeners. Consequently, some results of Experiment 2a have to be interpreted carefully based on potential differences across individual items, even though special care was taken for the statistical analysis (e.g. inclusion of random effects, inclusion of frequency) and some general patterns also seemed to emerge from my analysis. The following discussion will further link these results to previous literature and to my research questions.

5.1.3 Additional results and discussion

The results of Experiment 2a in Section 5.1.2 indicate that the duration of nasalization on vowels impacts the recognition of spoken words in L2 Canadian French. This suggests that English-native listeners are able to use vowel nasalization duration to distinguish between nasal ($\tilde{V}$) and coarticulated (VN) vowels in L2 French. Recognition patterns differed based on the proportion of nasalization duration (%NASDUR) values – when vowels were nasalized for a shorter period (20N or less), participants tended to identify them as coarticulated (CVN) more (i.e. 75% and 79% of trials) than when it was nasalized for a longer period of time (50N or more; less than 60% of trials). Although the difference was not as large as with L1 listeners in Experiment 1a, L2 listeners showed a tendency to distinguish between coarticulatorily nasalized and phonological nasal vowels. Furthermore, note that the effect of nasalization duration interacted with other factors (e.g. VOWEL QUALITY through TIME) in the eye movements statistical analysis, revealing complexities in the way L2 listeners process this information.

The ability to distinguish between short-nasalized (20N or less) and long-nasalized (50N or more) vowels is similar to what was found by Beddor et al. (2013) in L1 English. Recall that

Beddor et al. (2013) found that English listeners were sensitive to variations in nasalization duration in their L1. Consequently, the results from Experiment 2a can be taken as evidence that L2 listeners transferred knowledge from their L1 for L2 processing. As they are able to use coarticulatory nasalization in English to anticipate upcoming nasal consonants, they are also able to use fine-grained nasalization duration differences in L2 French. However, in French, they need to use this nasalization timing information to make a phonemic distinction between coarticulatory nasalized and phonological nasal vowels. In other words, for English speakers to learn French, they do not necessarily need to first learn to pay attention to French vowel nasalization as they are able to do that in English, but rather they need to establish a threshold of nasalization duration to differentiate between phonological nasal ($\tilde{V}$) and coarticulated nasalized (VN) vowels (among other factors, see Carignan 2014; Desmeules-Trudel & Brunelle, 2018).

One other interesting result of Experiment 2a concerns the potential failure of L2 listeners to *gradiently* perceive these small differences in %NASDUR values and apply them to a spoken word recognition task. Even if they were able to coarsely differentiate short-nasalized and long-nasalized vowels, it seems like within-category variability was not considered as much by L2 listeners than by L1 listeners in Experiment 1a, especially when they chose the nasal ($C\tilde{V}$) item. Keeping in mind potential across-item differences, the absence of a strong gradient pattern does not seem consistent with McMurray et al.'s (2002) gradient recognition hypothesis, even for moderate-to-advanced listeners of L2 French. However, if one assumes that a greater amount of cognitive resources is attributed to accessing lexical information in the L2 than in the L1 (for a review, see Clahsen & Felser, 2006), this leaves less available resources to access more fine-grained phonetic information. In other words, if the task of recognizing and identifying ambiguous word is harder in an L2 than in an L1, listeners might not pay as much attention to phonetic detail, and therefore rely on more generic knowledge (i.e. coarse-grained phonetic

information) about the L2 (e.g. short-nasalized vs. long-nasalized vowels, regardless of within-category variability). It is thus a possibility that given the difficulty of the task in the L2, listeners have stored fine-grained phonetic details but are not necessarily able to access it at all times.

Another possible explanation for the observed pattern is that L2 listeners, who do not have as much experience in the L2 as L1 listeners or who started learning L2 French at a later age, may not store phonetic details that are as fine-grained as in the L1. These differences between L1 and L2 processing can be attributed to the reduced amount of exposure in the L2, yielding less detailed representations in the L2, although results in the eye tracking paradigm are mixed concerning this factor of influence. If the fine-grained phonetic information is not present in the listeners' representations, it is not possible for them to access it. More research will be necessary to tease apart these two explanations for the observed eye movement patterns in Experiment 2a. For example, it would be interesting to investigate a contrast that exists in the L2 listeners' native language (e.g. [ɛ]-[a] L2 French words with native English listeners). As it is expected that L2 representations include fine-grained phonetic details for a contrast that exists in the L1, gradient vs. categorical patterns of recognition for [ɛ]-[a] words would enable me to determine if the failure to gradiently identify C $\tilde{V}$ -CVN words in the current experiment is due to increased cognitive load, or to the lack of stored phonetic details. More details on this future research will be presented in Section 6.4 below.

In general, the analysis of reactions times (RTs) revealed that for L2 listeners, the overall direction of the patterns regarding the increase or decrease in RTs and %NASDUR values was similar to L1 listeners in Experiment 1a (see Figure 3.6 compared to Figure 5.5). For example, as %NASDUR values increased, RTs to the recognition of phonological nasal (C $\tilde{V}$) words decreased, and inversely for the recognition of coarticulated nasalized (CVN) words (except for 0N

%NASDUR in 2 listeners). This suggests that the differences in processing load were relatively similar for L1 and L2 listeners.

Lastly, it is important to note that a number of additional effects were found to have a significant impact on L2 listeners' performance, especially concerning eye movement patterns. For example, I have suggested that familiarity (e.g. with *pomme*) and specific aspects of some stimuli could have influenced the GAMM predictions and significant terms. Furthermore, there was also a significant effect of TRIAL NUMBER (as well as interactions through TIME for eye movements) for the two eye movement GAMMs, which suggest that throughout the course of the experiment, listeners anticipated the stimuli. They "habituated" to the stimuli, in that they were able to more quickly decide to which image the heard stimulus corresponded. A number of FREQUENCY effects and interactions also came out significant of the statistical analyses (**Erreur ! Nous n'avons pas trouvé la source du renvoi.** e.g. Figure A.24 and Figure A.26), but are difficult to interpret given the small number of words that were used as critical items in the current experiment, and because the experimental design was not made to directly verify the influence of lexical frequency. I included it in the statistical models to account for as much variability in the data as possible, some which could be due to lexical frequency.

Note that as in Experiment 1a, the same stimuli in Experiment 2a were inherently ambiguous as some of them (50N, 80N and 100N) possessed conflicting cues: a long-nasalized vowel and a short nasal consonant. This could have influenced the results and led to effects that would not be found if stimuli were unambiguous. Natural connected speech in everyday life contains ambiguous word tokens, but isolated words do not vary as much. A second experiment, which was the same as Experiment 1b but with L2 listeners, was conducted with L2 speakers of CF to determine if the absence of conflicting phonetic cues had an influence on L2 word recognition, and to compare L2 listeners with L1 listeners in Experiment 1b above (Section 3.2).

5.2 EXPERIMENT 2B

Contrary to Experiment 2a, stimuli in Experiment 2b did not contain both a partly or fully nasalized vowel and a nasal consonant in the coda. Experiment 2b was a replication of Experiment 1b, but with L2 listeners instead of L1 listeners. This permitted me to test whether unambiguous words (i.e. that do not have conflicting phonetic cues) are processed differently than ambiguous ones (i.e. that have conflicting phonetic cues, as in Experiment 2a) in the L2.

5.2.1 *Methods*

5.2.1.1 *Participants*

Forty-seven L2 listeners of CF were paid or received partial course credit at the University of Ottawa for their participation. Forty-four were included in the current analyses (38 female, 6 male, between 17 and 29 years old, $\bar{X} = 18.9$ years, $SD = 2$), and 3 were excluded: one listener exceeded the 40-years-old age limit (63 years old), and two listeners spoke languages that contained phonological nasal vowels (one Tamil and one Polish). All listeners reported having normal hearing, normal or corrected-to-normal vision, and did not report any type or history of language, hearing or speech impairment. All participants completed a background language questionnaire (see Appendix 1), and self-reported knowing at least some French as an L2 at a low to high level of proficiency, as well as other languages. Importantly, no speaker that is included in the analyses had been exposed to another language containing contrastive nasal vowels in its phonological inventory. Furthermore, 16 listeners were early bilinguals who were exposed to L2 French in a naturalistic setting, and 28 listeners were “late” bilinguals who were exposed to French in a classroom context, according to the same criteria as in Experiment 2a. The

distribution of self-reported L2 French proficiency scores in the early bilinguals and late bilinguals groups is presented in Figure 5.16.

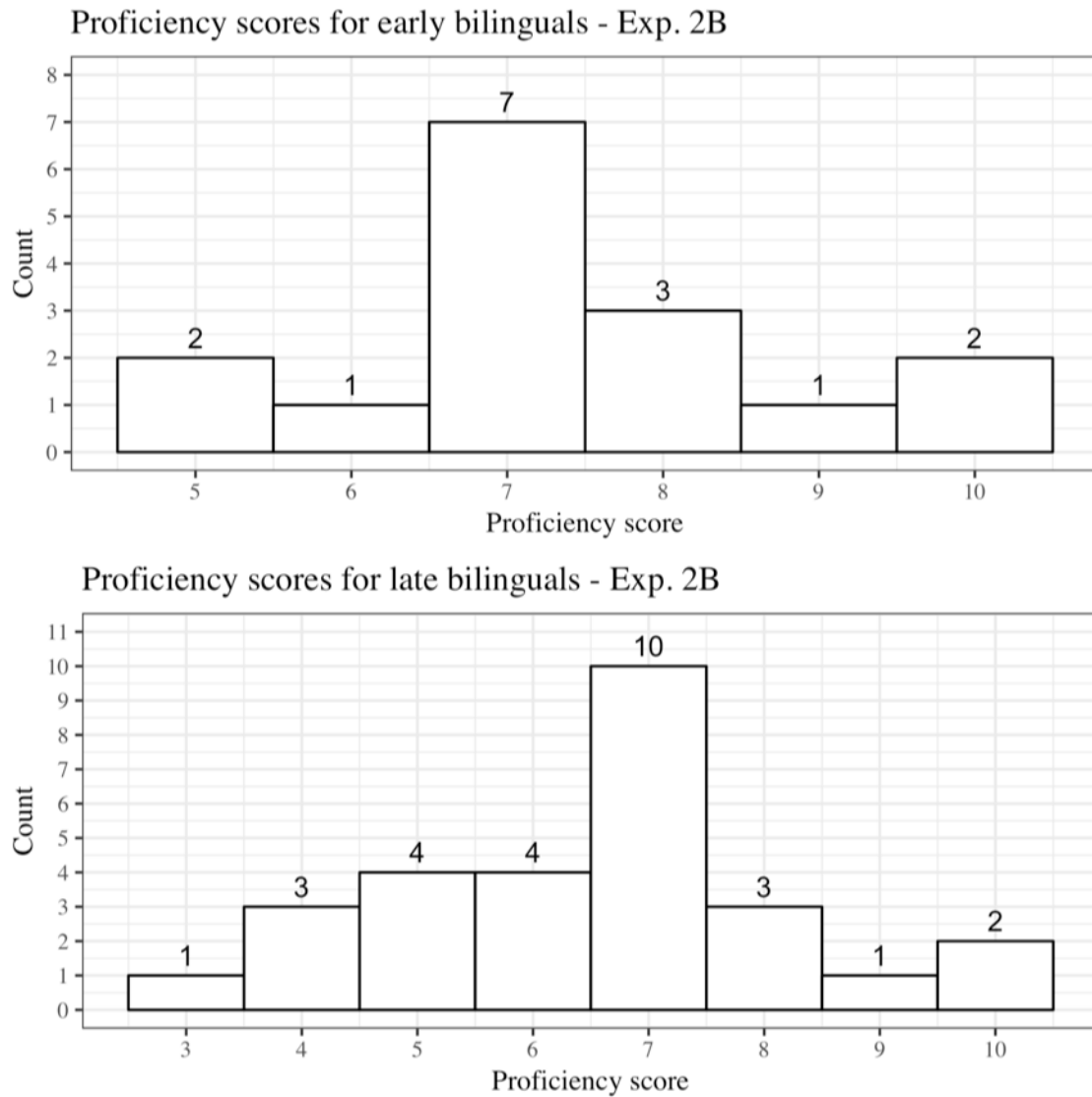


Figure 5.16. Distribution of self-reported L2 French proficiency scores out of 10 for Exp. 2b participants

The average self-reported current exposure to French as a L2 for the participants of Experiment 2b is also presented in Figure 5.17, suggesting that school and friendships are the two contexts in which the groups were exposed to and used French as a L2 the most.

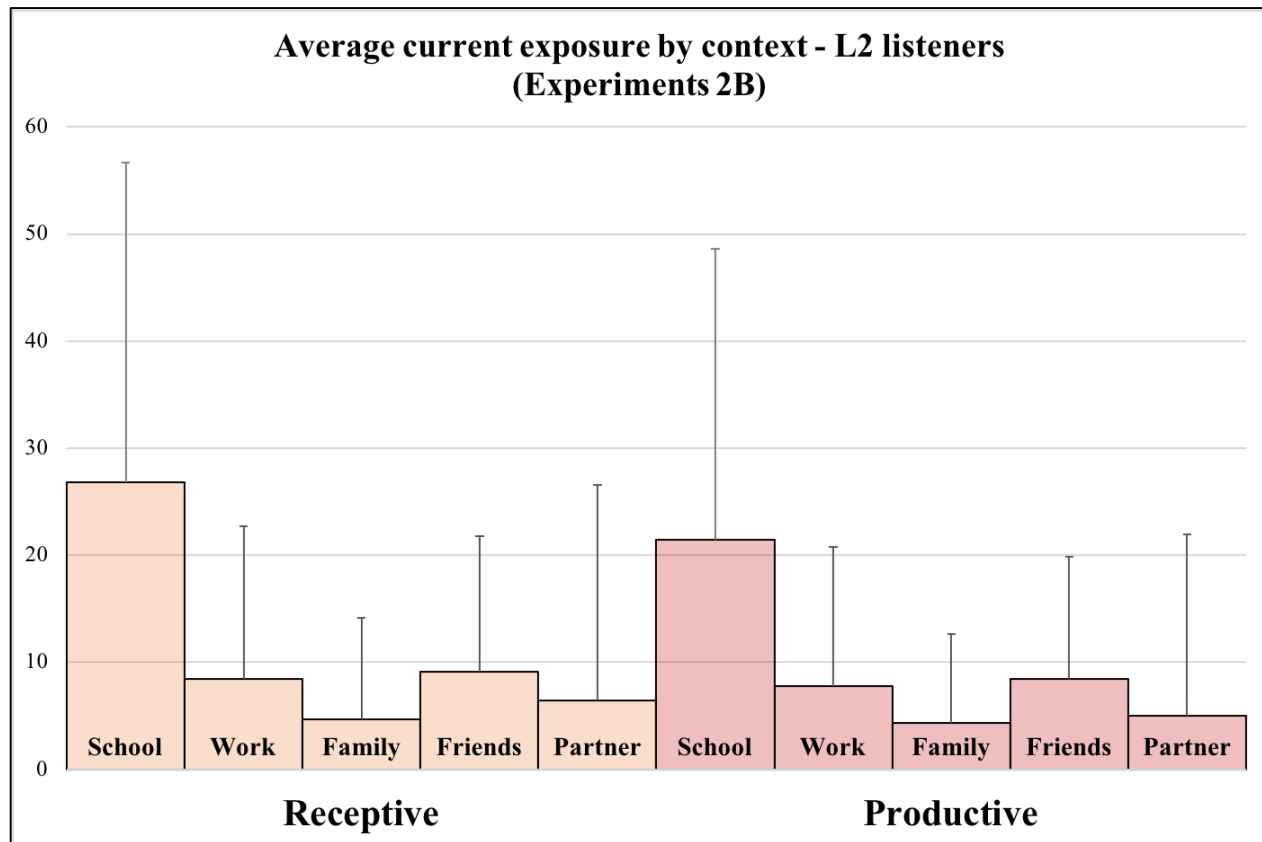


Figure 5.17. Average proportion of current exposure (receptive) to and use (productive) of L2 French in various situations for Exp. 2b participants

5.2.1.2 Stimuli

Word triads were the same as in Experiment 2a (and experiments with L1 listeners; see Table 5.2). They were paired with the same images used in the previous experiment. The auditory stimuli were the same as in Experiment 1b (see Section 3.2.1.2), i.e. stimuli that did not contain a

50-ms nasal appendix (Figure 3.15 in Section 3.2.1.2), but did vary in duration of nasalization. The same duration of nasalization conditions (0N, 20N, 50N, 80N and 100N) were used.

5.2.1.3 Procedure and analysis

The experimental procedure, including calibration, familiarization, stimuli presentation, and collected data was the same as Experiments 2a. The analysis (GAMMs) and investigated variables were the same as Experiment 2a, and tables summarizing the statistical predictors are presented just before presenting the results.

5.2.2 Results

This section presents the results of Experiment 2b, focusing on the chosen images and reaction times (RTs) in Section 5.2.2.1. The eye tracking results are presented in Section 5.2.2.2. Recall that the chosen image measure presented below corresponds to the image that was fixated the longest in the last 1000 ms of each trial.

5.2.2.1 Chosen image and reaction times

Figure 5.19 shows that in Experiment 2b, L2 listeners chose the nasal (C $\tilde{V}$) image to a greater proportion (between 36.07% and 79.53%) than the coarticulated (CVN) image (between 17.97% and 26.35%; upper left panel). Also note that in the 0N condition, participants chose the oral (CVC) image 42.09% of the time. Similarly to L1 listeners in Experiment 1b (Section 3.2.2.1), the increased rate of oral (CVC) word choices is due to the cross-splicing procedure. In the 0N condition, the vowel from a CVC item was pasted within the CVN matrix, but the formant transitions at the end of the vowel led to a following oral consonant with a place of articulation

that was sometimes different from the nasal consonant in the CVN word (e.g. *dame* [dam] ‘lady’ ~ *datte* [dat] ‘date (fruit)’). As the nasal consonant from the CVN word was fully removed, participants who were presented with auditory stimuli containing transitions towards the oral consonant chose the oral image as the best fit in this condition. A number of trials in which listeners reported not knowing one of the members of the pair in the familiarization (i.e. 750/3960) were rejected for analysis. Of the remaining 3210 trials, 2609 C $\tilde{V}$ and CVN image choices were included in the analysis (also see Figure 5.22). The results presentation and GAMM analysis (bottom panel in Figure 5.18 and Table 5.9) will focus on the nasal (C $\tilde{V}$) and coarticulated (CVN) items, as these are the two main items of interest for the question of the influence of nasalization duration and the recognition of fine-grained variations of phonetic cues.

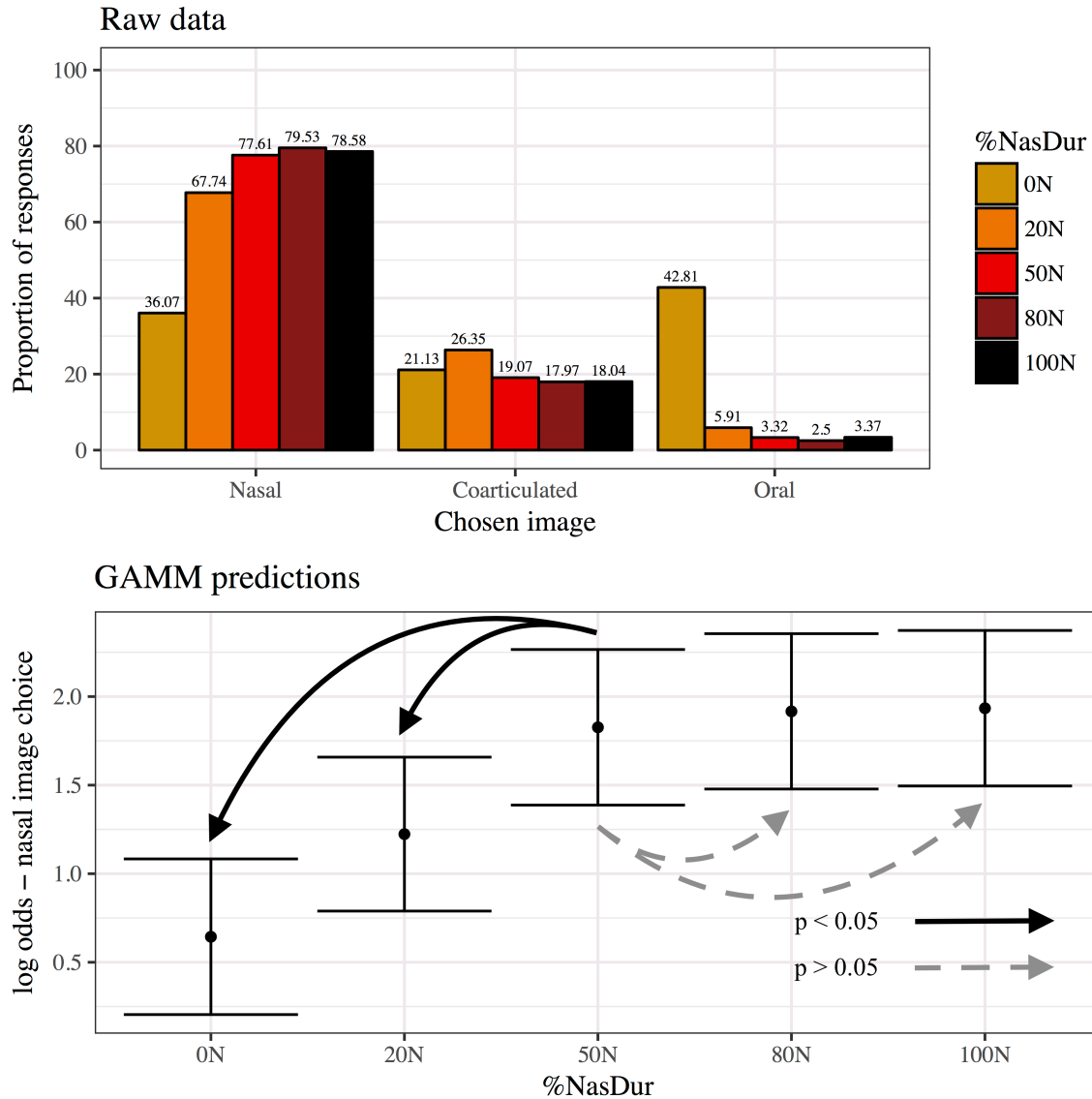


Figure 5.18. Proportions of image choices for nasal (C $\tilde{V}$) and coarticulated (CVN) images (upper panel) and GAMM predictions for image choice analysis (i.e. probability of C $\tilde{V}$ image choice by %NASDUR) – Exp. 2b

The GAMM analysis suggests that when compared to the 50N baseline, the probability of choosing the C $\tilde{V}$ image in the 50N %NASDUR condition is significantly different (see Table 5.9) from the 0N and 20N levels of %NASDUR, but not from 80N and 100N. This suggests that the probability of choosing the C $\tilde{V}$ image gradually increased between 0N and 50N %NASDUR, and remained stable across 50N, 80N and 100N nasalization duration. This suggests a categorical

pattern of perception, where within-category variability between 50% and 100% nasalization duration did not impact recognition, which can be interpreted as a categorical recognition effect. Additional effects are illustrated in Appendix 3.

Table 5.9

GAMM results – Exp. 2b image choices

A. parametric coefficients	Estimate	Std. Error	t-value	p-value	
Intercept	1.8320	0.3953	4.6344	< 0.0001	***
%NasDur[0N]	-1.1970	0.1835	-6.5232	< 0.0001	***
%NasDur[20N]	-0.6080	0.1607	-3.7822	0.0002	***
%NasDur[80N]	0.0822	0.1688	0.4873	0.6261	
%NasDur[100N]	0.1034	0.1696	0.6098	0.5420	
B. smooth terms	edf	Ref.df	F-value	p-value	
s(L2_proficiency)	2.7502	2.9076	27.1380	< 0.0001	***
Random_Participant	33.0424	42.0000	229.4119	< 0.0001	***
Random_Item	7.7433	8.0000	278.4736	< 0.0001	***

Concerning RTs in Figure 5.19, for both nasal (C $\tilde{V}$) and coarticulated (CVN) word choices, 0N %NASDUR yielded the largest RT values on average. The 20N %NASDUR yielded intermediate values between 0N and 50N, and the other three %NASDUR led to shorter and similar RTs. This is consistent with the fact that listeners chose the C $\tilde{V}$ image in a greater proportion overall, suggesting that processing load for auditory stimuli that were nasalized for 50% or more of their duration as nasal C $\tilde{V}$ words was lower than when they were nasalized for a shorter period. Shorter average RTs for the nasal C $\tilde{V}$ image, as opposed to the coarticulated CVN image, also suggests lighter processing load when participants categorized the stimuli as C $\tilde{V}$.

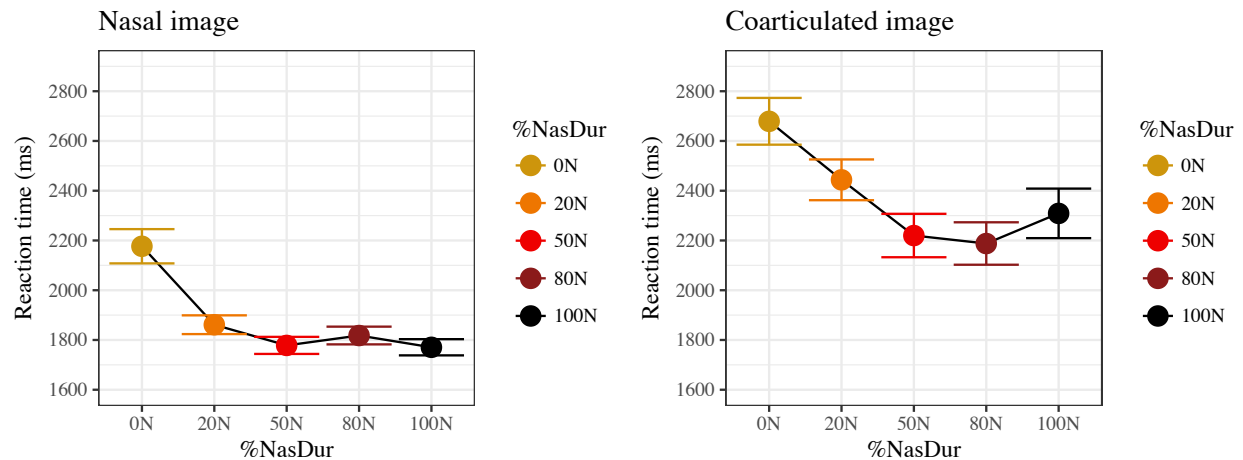


Figure 5.19. RT averages for nasal (C $\tilde{V}$) image choices by %NASDUR (left) and coarticulated (CVN) image choices (right) by %NASDUR – Exp. 2b

The remainder of this section presents the results of the GAMM analysis of RTs, using the maximum likelihood (ML) method (for a similar model fitting procedure, see Porretta et al., 2016). The input variable of the model was the log-transformed RT. Random effects structure was determined by maximum-likelihood chi-square tests (*compareML()* function from the *itsadug* package), therefore random smooth (i.e. nonlinear) terms by PARTICIPANTS and %NASDUR, and by ITEM (minimal pair label) and %NASDUR were kept in. Predictors of interest and interactions are summarized in Table 5.10.

Table 5.10

Summary of the independent variables for the full RT GAMM – Exp. 2b

Parametric terms	Levels
CHOSEN IMAGE	Factor w/ 2 levels: Nasal, Coarticulated image
VOWEL QUALITY	Factor w/ 3 levels: [a], [e], [o]
EXPOSURE TO L2 BEFORE 5	Logical w/ 2 levels: True or False
Smooth terms	Levels
%NASDUR	Numerical (centered): -50 to 50
TRIAL NUMBER	Numerical: 1 to 160
LOG FREQUENCY (CVN)	Numerical: -1.0804 to 1.1756
LOG FREQUENCY (C $\hat{V}$)	Numerical: -0.5229 to 0.7491
L2 FRENCH PROFICIENCY	Numerical: 3 to 10
Interactions	
%NASDUR*VOWEL QUALITY	
%NASDUR*FIXATED IA	
Random effects	
Smooth term:	
PARTICIPANTS*%NASDUR	
Smooth term:	
MINIMAL PAIR LABEL*%NASDUR	

A backwards stepwise procedure for GAMM fitting was used, as in the previous experiments. Results of the final model are presented in Table 5.11 with CHOSEN IMAGE and VOWEL QUALITY as parametric factors, and significant smooth terms for %NASDUR, TRIAL, and the interaction between %NASDUR and VOWEL QUALITY. Note that p -values lower than 0.05 here suggest that the smooth term is significantly different from zero, suggesting statistical difference between two or more points in the continuum (e.g. %NASDUR or TRIAL NUMBER). Presentation of the results will focus on the significant %NASDUR smooth term and %NASDURXVOWEL QUALITY interaction, which are crucial to the research question. Illustration of the additional significant factor (TRIAL) is found in Appendix 3.

Table 5.11

GAMM results – Exp. 2b RTs

A. parametric coefficients	Estimate	Std. Error	t-value	p-value
Intercept	3.3006	0.0318	103.9038	< 0.0001
Chosen_image	-0.0501	0.0068	-7.3329	< 0.0001
Vowel_[e]	-0.0174	0.0384	-0.4516	0.6516
Vowel_[o]	0.0013	0.0385	0.0337	0.9731
B. smooth terms	edf	Ref.df	F-value	p-value
s(%NasDur)	2.7211	2.8770	12.7046	< 0.0001
s(Trial)	2.8287	3.4470	3.8745	0.0070
s(%NasDur):Vowel_[a]	2.7606	2.9110	4.3269	0.0028
s(%NasDur):Vowel_[e]	1.0002	1.0003	4.1036	0.0429
s(%NasDur):Vowel_[o]	0.0008	0.0014	0.1086	0.9902
Random_Participant	67.1407	174.0000	8.3864	< 0.0001
Random_Item	12.6811	30.0000	7.0235	< 0.0001

Parametric coefficient estimates in Table 5.11 suggest that overall, when participants chose the nasal (C $\tilde{V}$) image, RTs were significantly different from when participants chose the coarticulated (CVN) image (3.301 ± 0.0501 , $t = -7.333$, $p < 0.05$). This matches the raw RT values presented in Figure 5.19. The parametric VOWEL QUALITY factor, although not significant, was kept into this section of the model in order to offer an intercept value for the significant smooth interaction that includes VOWEL QUALITY. Figure 5.20 presents the data that is predicted by the model according to the %NASDUR smooth term.

Figure 5.20 presents the data that is predicted by the GAMM concerning the effect of %NASDUR. It shows that when vowels are not nasalized (in the 0N %NASDUR), participants displayed the longest RTs compared to other %NASDUR values. When the vowel was nasalized for a short period only (20N), it is shown that RT values are intermediate between 0N and 50N, and when the vowel was nasalized for half or more of its duration (50N, 80N, and 100N), RTs seem overall similar, perhaps slightly longer in the 80N %NASDUR. This seems consistent with

the fact that overall, stimuli that corresponded to the canonical realization of CF phonological nasal vowels ($\tilde{V}$; that are nasalized for 50% or more of their duration) led to lower processing load, as opposed to stimuli which did not correspond (0N and 20N) to the canonical realization of either phonological nasal ($\tilde{V}$) or coarticulated nasalized (VN) vowels and incurred longer RTs.

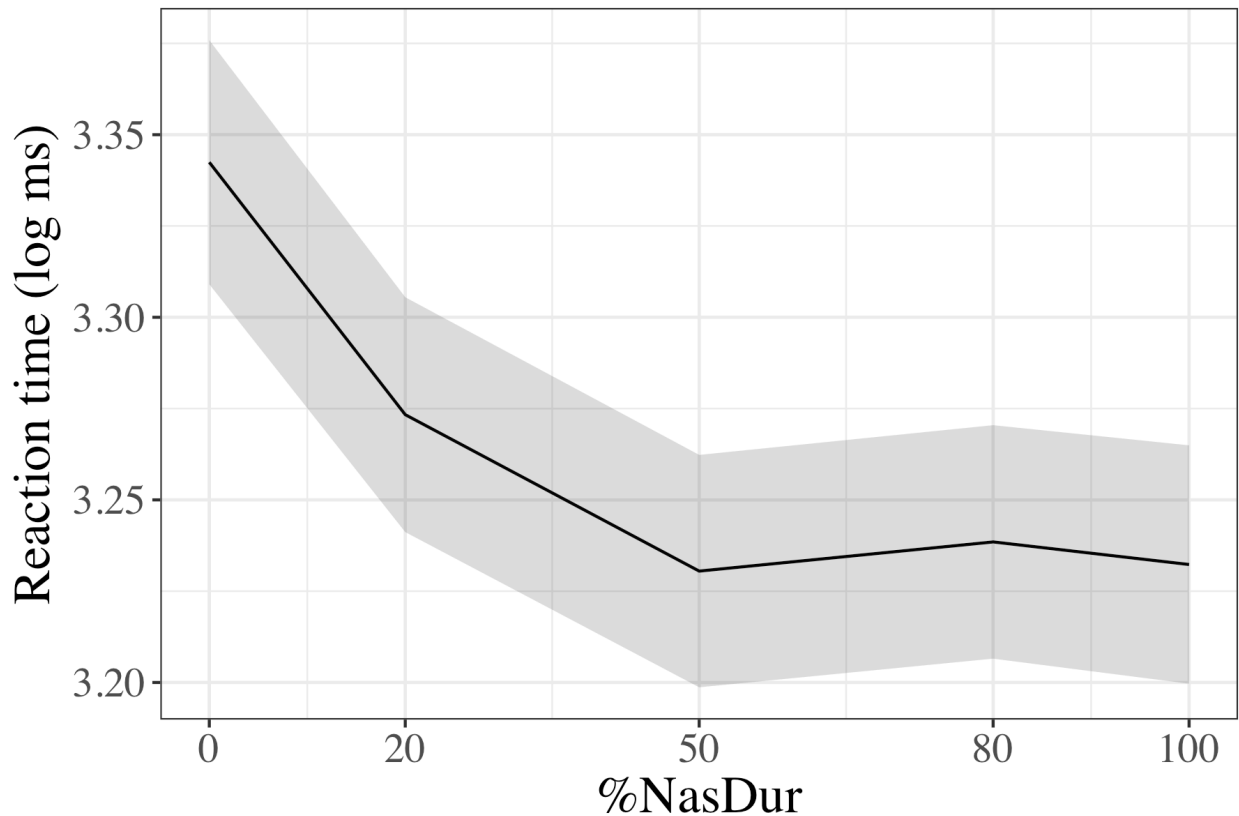


Figure 5.20. Predicted data by %NASDUR – Exp. 2b RTs

The predicted data in Figure 5.21 shows the effect of the %NASDUR and VOWEL QUALITY interaction on RTs. In the 0N %NASDUR condition, [a-ã] yield longer RTs on average than [ɔ-õ], which in turn yield longer RTs than [ɛ-ẽ]. In the 20N %NASDUR, [a-ã] words yield shorter RTs than [ɔ-õ] words, and [ɛ-ẽ] words still yield shorter RTs. For the other three %NASDUR values, RTs seem equivalent across VOWEL QUALITIES. Despite some differences when the stimuli vowels

were nasalized for a short period of time, which could be due to individual items, patterns of decrease as %NASDUR values increase seem.

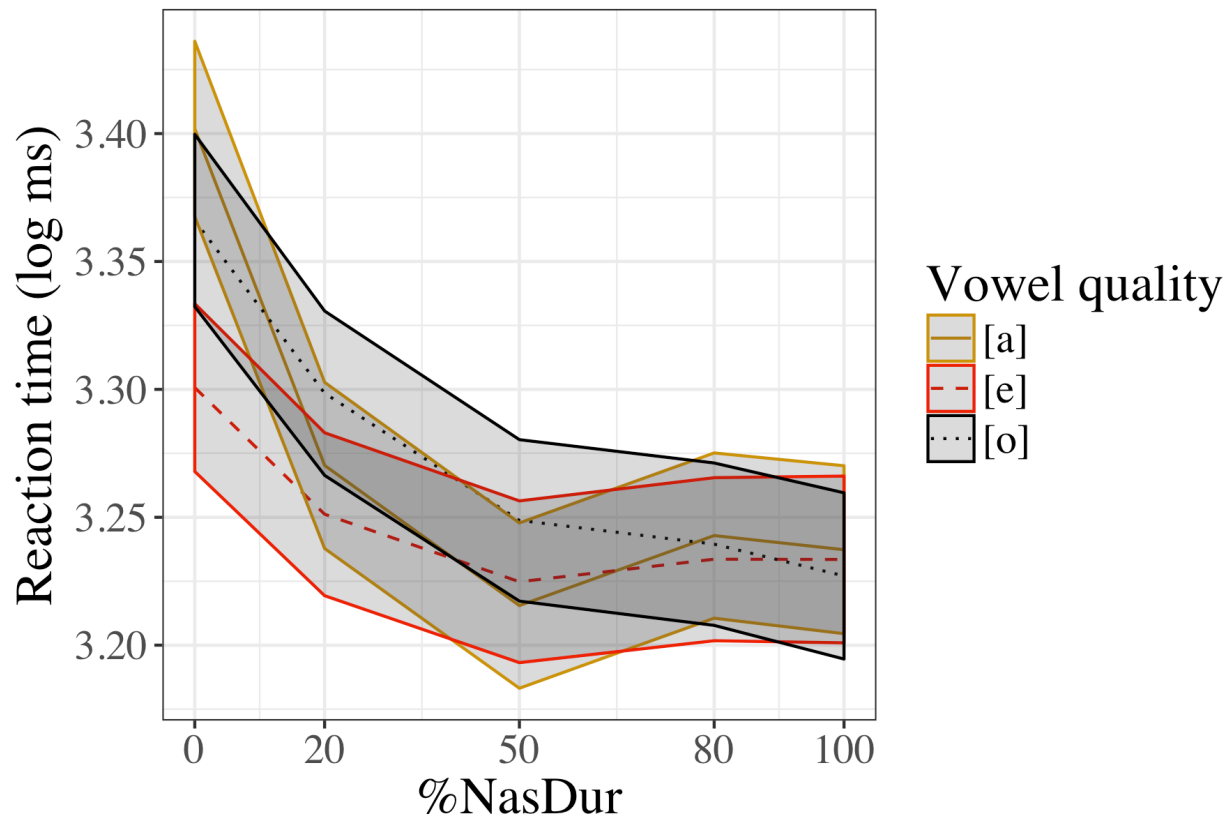


Figure 5.21. Predicted data by %NASDUR and VOWEL QUALITY – Exp. 2b RTs

5.2.2.2 Eye movements

In Experiment 2b, participants chose the C $\tilde{V}$ image 2008 times, and chose the CVN image 601 times. Figure 5.22 presents patterns of fixations to the C $\tilde{V}$ (leftmost panels) and CVN images (rightmost panels) when listeners chose the C $\tilde{V}$ image (upper panels) and when they chose the CVN image (lower panels) in the 200-950 ms time window. Observing fixation patterns in the upper left panel of Figure 5.22 (fixations to C $\tilde{V}$, C $\tilde{V}$ choices), it seems like the 50N-to-100N %NASDUR conditions behave together, and 20N and 0N are different. On the other hand, when

listeners chose the coarticulated (CVN) image, the 0N and 100N %NASDUR behaves differently from the other conditions, especially at the end of the time window. The other three %NASDUR values display similar patterns, except for 50N at the beginning of the time window. This seems generally consistent with the image choice analysis in 5.2.2.1.

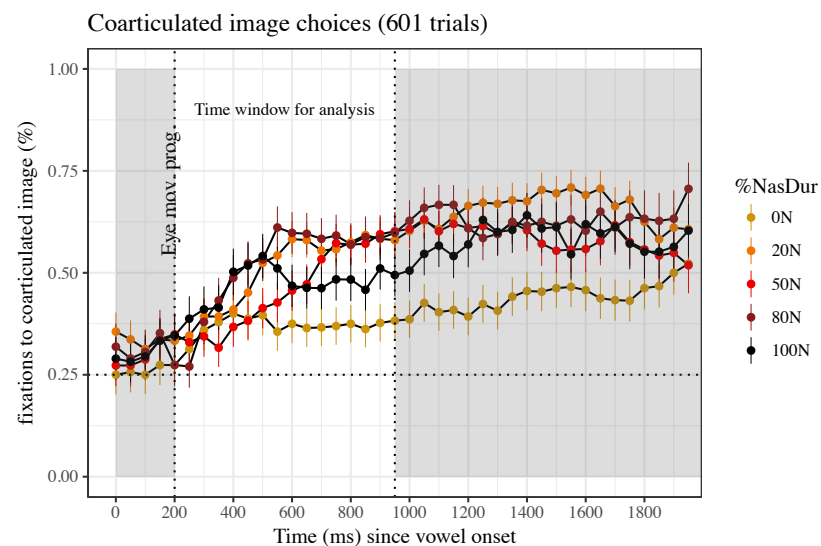
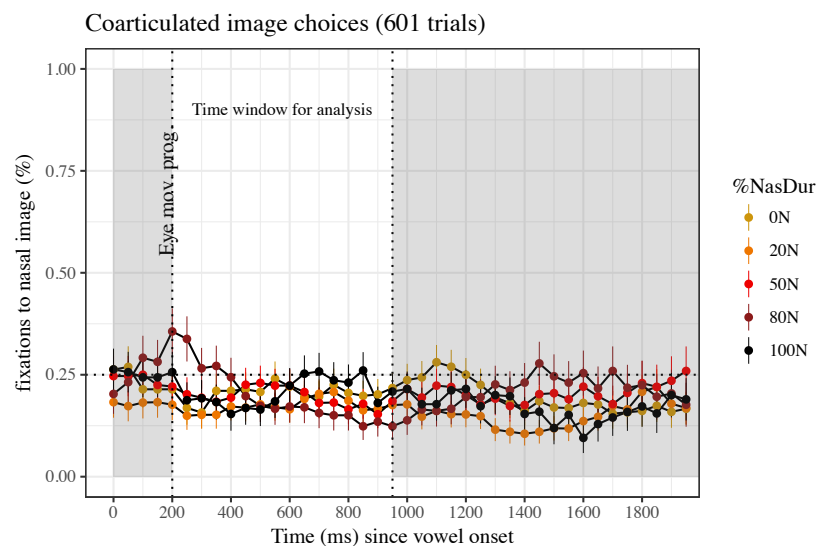
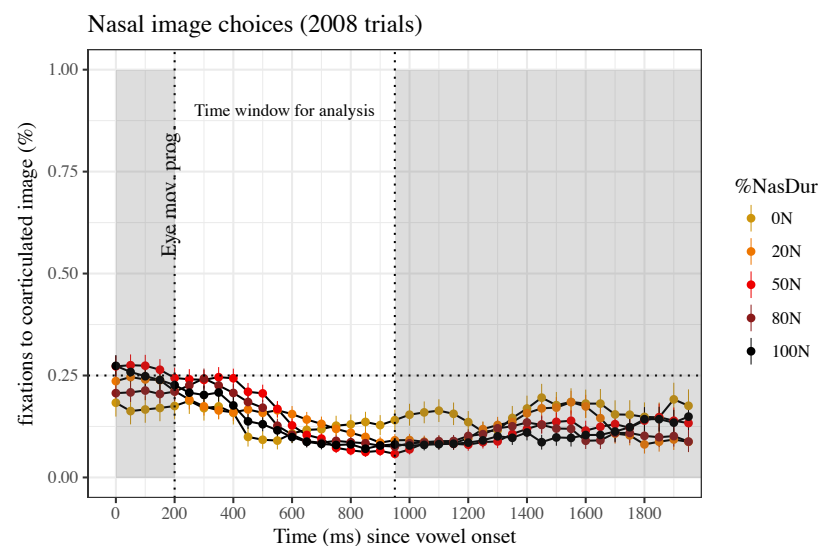
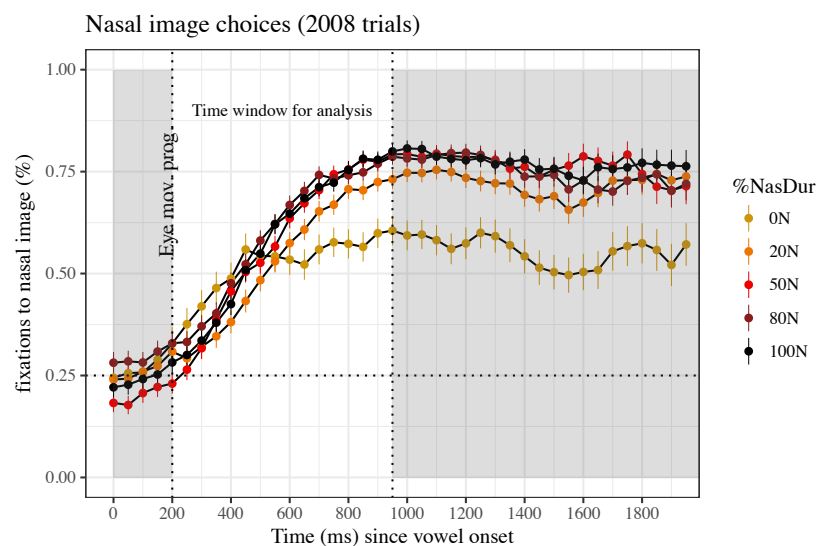


Figure 5.22. Proportions of fixations to the nasal ($C\tilde{V}$) image (leftmost) and coarticulated (CVN) image (rightmost), nasal ($C\tilde{V}$) choices (upper) and coarticulated (CVN) choices (lower) – Exp. 2b

The models that were fitted to the eye-tracking data used the maximum likelihood (ML) method (for a similar model fitting procedure, see Porretta et al., 2016) and a stepwise backwards fitting procedure. The input variable of the GAMM model was the empirical logit-transformed fixations (Barr, 2008) to the nasal (C $\tilde{V}$) image in the first model, and to the coarticulated (CVN) image in the second model. The predictors of interest that were assessed in the statistical analysis are summarized in Table 5.12. Note that they are the same as in Experiment 1b, except for the PROVINCE OF ORIGIN factor which was not considered in Experiment 2b, and the L2 FRENCH PROFICIENCY factor which was added in, similarly to Experiment 2a. The low number of participants for some levels of the PROVINCE factor (only one participant was from Nova Scotia, and one from Saskatchewan) would not enable me to reliably interpret the results of the statistical outcome for this variable. Furthermore, it does not directly contribute to answering the question of the capacity of L2 French listeners to integrate nasalization duration during spoken word recognition.

Table 5.12

Summary of the independent variable for the full GAMM – Exp. 2b %fix

Parametric terms	Levels
CHOSEN IMAGE	Factor w/ 2 levels: Nasal (baseline), Coarticulated image
VOWEL QUALITY	Factor w/ 3 levels: [a], [e], [o]
EXPOSURE TO L2 BEFORE 5	Logical w/ 2 levels: True or False
Smooth terms	Levels
TIME	Numerical: 0 to 750, separated in 50 ms time bins
%NASDUR	Numerical (centered): -50 to 50
TRIAL NUMBER	Numerical: 1 to 160
LOG FREQUENCY (COARTICULATED)	Numerical: -1.0804 to 1.1756
LOG FREQUENCY (NASAL)	Numerical: -0.5229 to 0.7491
L2 FRENCH PROFICIENCY	Numerical: 7 to 10
Interactions	
TIME* FIXATED IA	
TIME* VOWEL QUALITY	
TIME* PROVINCE OF ORIGIN	
TIME* EXPOSURE TO L2 BEFORE 5	
TIME*%NASDUR	
TIME* TRIAL NUMBER	
TIME* LOG FREQUENCY (COART.)	
TIME* LOG FREQUENCY (NASAL)	
TIME* L2 FRENCH PROFICIENCY	
TIME*%NASDUR *VOWEL QUALITY	
TIME*CONDITION*FIXATED IA	
Random effects	
Smooth term: PARTICIPANTS*%NASDUR	
Smooth term:	
MINIMAL PAIR LABEL*%NASDUR	

Results of the model on proportions of fixations to the C $\tilde{V}$ item are presented in Table 5.13 with VOWEL QUALITY (as a reference intercept for the interaction in the nonlinear part of the GAMM), significant smooth terms for TIME, TRIAL, %NASDUR, and FREQUENCY (COARTICULATED) as a baseline for the higher-order interaction, as well as 2-way interactions between TIME and VOWEL QUALITY, and TRIAL NUMBER, and a 3-way interaction between TIME, CONDITION, and CHOSEN IMAGE.

Table 5.13

GAMM results – Exp. 2b %fix to C $\tilde{V}$, C $\tilde{V}$ choices

A. parametric coefficients	Estimate	Std. Error	t-value	p-value	
Intercept	0.6800	0.3372	2.0167	0.0437	*
Vowel_[e]	0.2274	0.4598	0.4945	0.6209	
Vowel_[o]	-0.2908	0.4341	-0.6699	0.5030	
B. smooth terms	edf	Ref.df	F-value	p-value	
s(Time)	4.8266	5.3045	6.5893	< 0.0001	***
s(Trial)	2.6906	3.3450	5.0952	0.0012	**
s(%NasDur)	2.0988	2.4769	17.3447	< 0.0001	***
s(Freq_coart.)	1.0081	1.0091	0.4246	0.5141	
s(Time):Vowel_[a]	0.0075	0.0089	0.0668	0.9805	
s(Time):Vowel_[e]	2.4174	2.6630	1.2276	0.2726	
s(Time):Vowel_[o]	4.0692	4.5402	2.1953	0.0570	
s(Time: Trial)	7.2518	9.8740	2.7606	0.0023	**
s(Time:%NasDur)	11.3344	13.5906	3.7729	< 0.0001	***
s(Time:Freq_coart.)	1.5594	1.6434	1.9523	0.0925	
s(Time:%NasDur):Vowel_[a]	0.0069	0.0112	0.0482	0.9814	
s(Time:%NasDur):Vowel_[e]	6.4436	8.9218	0.7139	0.6761	
s(Time:%NasDur):Vowel_[o]	4.3251	6.2216	2.7700	0.0111	*
Random_Participant	195.8764	377.0000	1.7955	< 0.0001	***
Random_Item	50.3945	77.0000	4.3496	< 0.0001	***

Presentation of the results will focus on the significant %NasDur factor, and the timeXvowel quality ($p = 0.057$), timeX%NasDur, and timeX%NasDurXchosen image interactions. Illustrations of the additional significant smooths (TRIAL NUMBER, and interactions between TIME and TRIAL, and FREQUENCY (COARTICULATED)) are found in Appendix 3. Each smooth term is detailed and illustrated below. Auto-correlation AR1 value, corresponding to the average correlation between one data point and the preceding one in the time series, was empirically set to 0.818 based on the data.

In Figure 5.23, it is shown that %NASDUR has a significant influence on proportions of fixations (elog-transformed) to the nasal (C $\tilde{V}$) image. When the vowel is nasalized between 0% and 80% of its duration, there seems to be a constant increase in fixations. There then seems to be

a stabilization in proportions of fixations between 80% and 100% nasalization duration. Based on the illustration of raw proportions of fixation through time in Figure 5.22 **Erreur ! Nous n'avons pas trouvé la source du renvoi.**, this rising pattern across %NASDUR values is expected.

However, note that the time factor is not considered here, and will be further detailed below (see Figure 5.26).

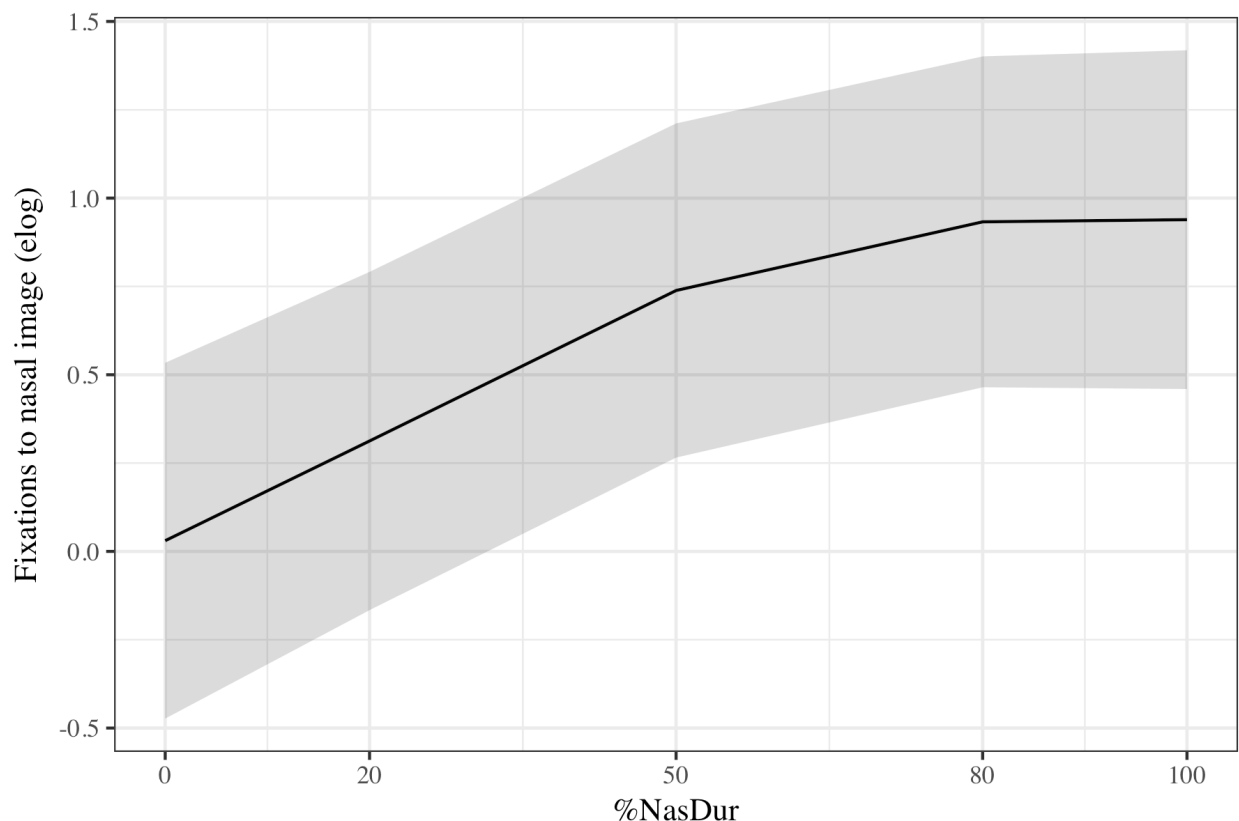


Figure 5.23. Predicted data by %NASDUR – Exp. 2b elog %fix to $C\tilde{V}$, $C\tilde{V}$ choices

In Figure 5.24, it is shown that at the beginning of the time window until 400 ms, stimuli with the [ɔ-ɔ̃] vowels yielded slightly less fixations to the nasal ($C\tilde{V}$) image than the other VOWEL QUALITIES. However, at the end of the window, fixations are similar across vowel qualities. This discrepancy is consistent with the finding of Experiment 1b that in the 0N-20N %NASDUR, the

[ɔ̃-ɔ̃] stimuli led to higher proportions of fixations to the CVN target. This effect was likely due to the behaviour of *rond-rhum*. Furthermore, in this condition, back rounded [ɔ̃] was at the end of the word, which is phonotactically illegal in French (e.g. ‘Caroline (surname)’ /kaʁɔ̃lin/ → ‘Caro (nickname)’ [kaʁɔ̃]). This additional constraint on the realization of [ɔ̃-ɔ̃] stimuli in 0N %NASDUR could explain the different behaviour of [ɔ̃-ɔ̃] words in Experiment 2b.

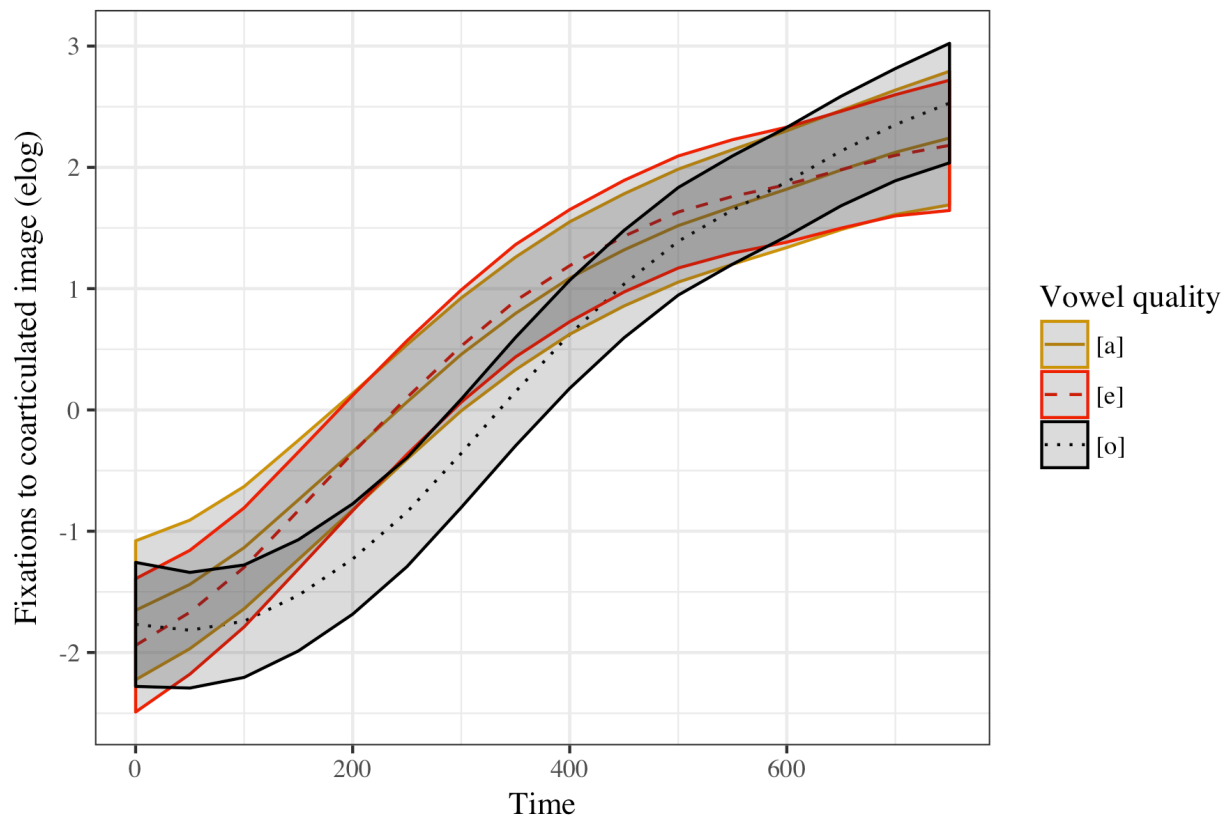


Figure 5.24. Predicted data by VOWEL QUALITY through TIME – Exp. 2b elog %fix to C $\tilde{V}$, C $\tilde{V}$ choices

Figure 5.25 shows the effect of %NASDUR values on fixations to the nasal (C $\tilde{V}$) image through TIME, offering a detailed picture of the effect that was explained in Figure 5.23 above. In the 0N %NASDUR value (golden-coloured solid line), fixations to the nasal (C $\tilde{V}$) image remain

low throughout the time window. The 20N %NasDur value also led to slightly smaller proportions of fixations, however, it still followed a similar pattern to the other %NasDur conditions. The 50N, 80N and 100N seem overall similar, suggesting a more categorical pattern of word recognition similar to L1 listeners in Experiment 1b above.

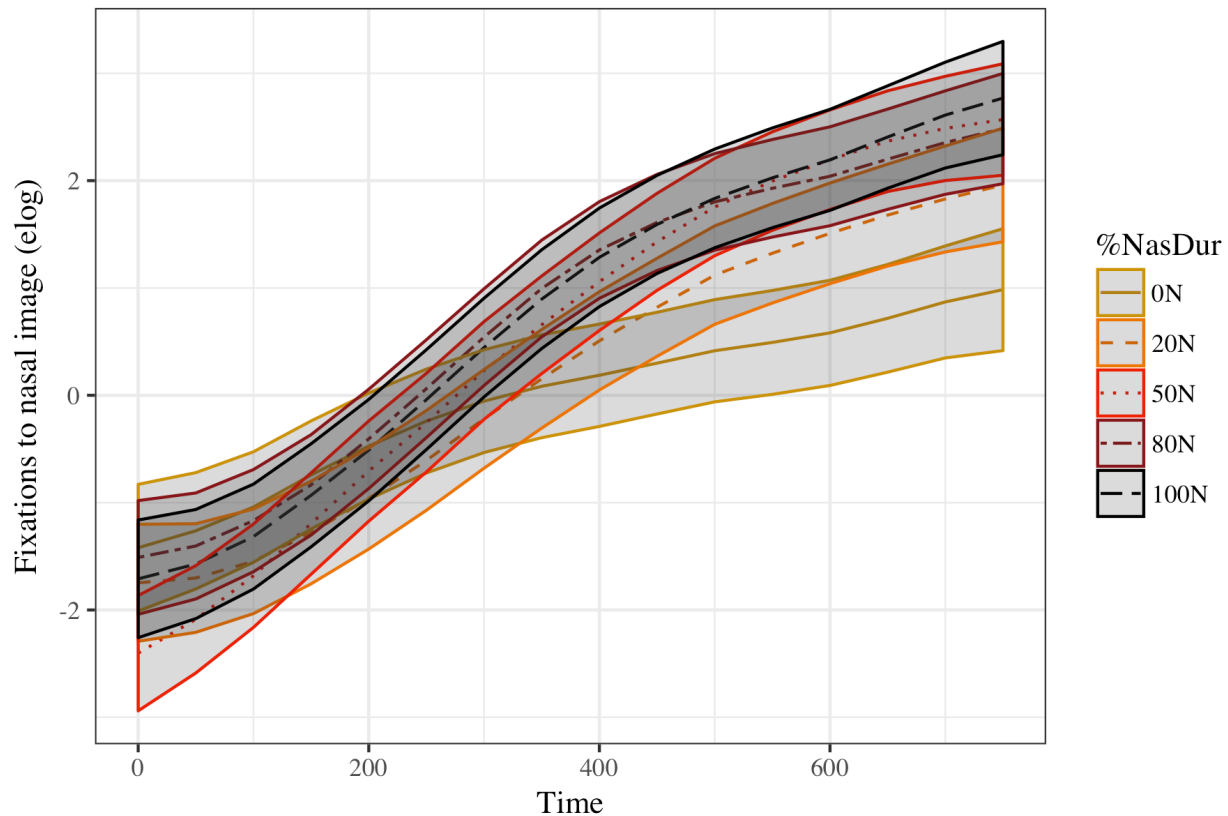


Figure 5.25. Predicted data by CHOSEN IMAGE through TIME – Exp. 2b elog %fix to C $\tilde{V}$, C $\tilde{V}$ choices

Figure 5.26, similarly to all other eye tracking analyses, illustrates the significant interaction between %NasDur and vowel quality through time. Again, it is apparent that patterns of fixations to the C $\tilde{V}$ image when participants chose this image were complex. Mainly, it seems like the 0N %NasDur yielded the lowest proportions of fixation through time to the C $\tilde{V}$ image for

[o] stimuli, higher proportions for [a], and higher proportions for [e] stimuli. This could be due to the familiar *pomme* item, which would have conditioned the participants to fixate it more than its nasal counterpart *pont*, and therefore have caused lower proportions of fixations to the C $\tilde{V}$ item.

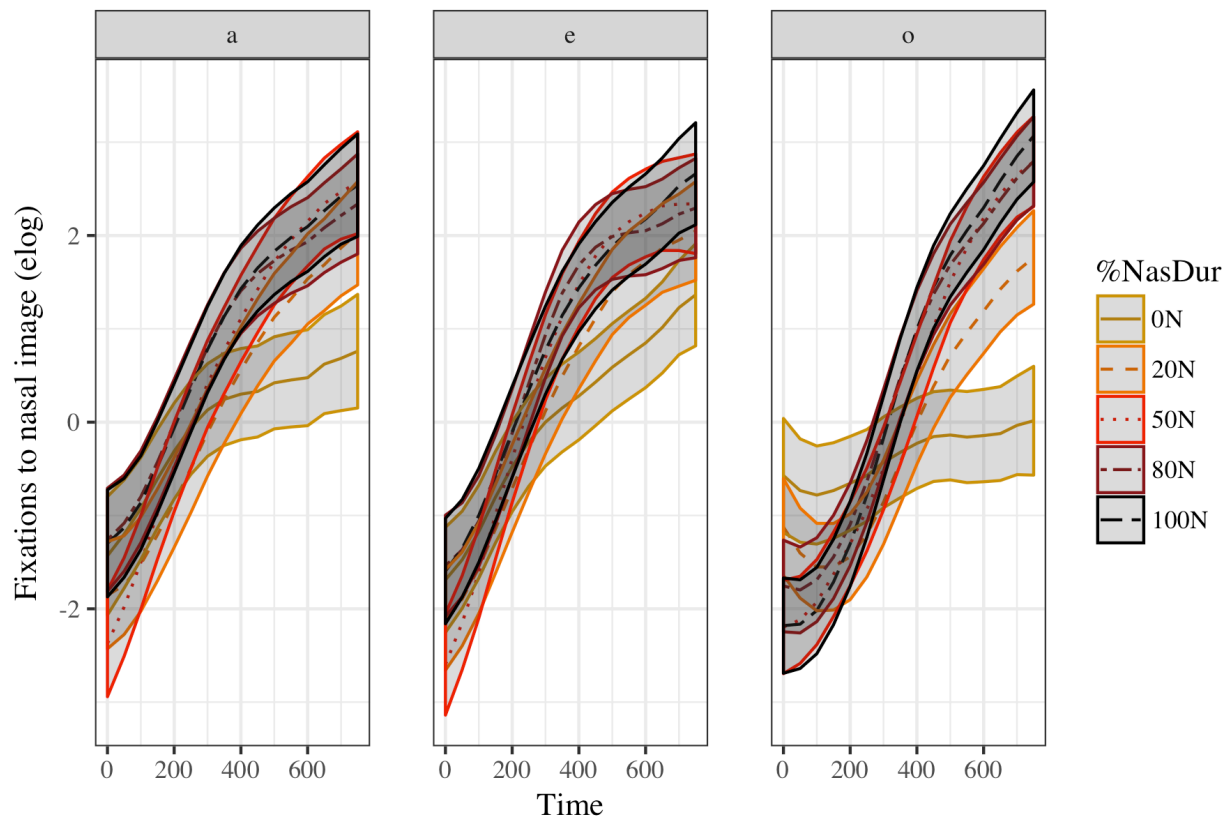


Figure 5.26. Predicted data by %NASDUR and VOWEL QUALITY through TIME – Exp. 2b elog %fix to C $\tilde{V}$, C $\tilde{V}$ choices

The second GAMM for Experiment 2b, which models the (empirical-logit transformed) proportions of fixations to the coarticulated (CVN) image when participants chose the CVN image (601 trials) is presented in Table 5.14. In the main text, I will focus on the %NASDUR smooth term, as well as the TIMEX%NASDURXVOWEL QUALITY interaction. Note that the non-significant TIMEX%NASDUR interaction remained in the model as a baseline to the higher-order

interaction. The other significant smooths and interactions (TIME, FREQUENCY (COARTICULATED), TIME $\times$ TRIAL, TIME $\times$ FREQUENCY (COARTICULATED)) are illustrated in Appendix 3.

Table 5.14

GAMM results – Exp. 2b %fix to CVN, CVN choices

A. parametric coefficients	Estimate	Std. Error	t-value	p-value	
Intercept	-1.0982	0.2844	-3.8616	0.0001	***
Vowel_[e]	1.2541	0.3311	3.7872	0.0002	***
Vowel_[o]	1.0287	0.2909	3.5368	0.0004	***
B. smooth terms	edf	Ref.df	F-value	p-value	
s(Time)	2.8168	3.2745	7.8239	< 0.0001	***
s(Trial)	2.6100	3.2289	2.2531	0.0800	
s(%NasDur)	2.5112	2.8157	3.5212	0.0102	*
s(Freq_coart.)	2.0230	2.3401	5.5763	0.0038	**
s(Time:Trial)	12.5907	14.6923	3.0189	< 0.0001	***
s(Time:%NasDur)	9.8816	11.9211	1.2346	0.2917	
s(Time:Freq_coart.)	6.7721	8.5420	2.0125	0.0330	*
s(Time:%NasDur):Vowel_[a]	7.2969	9.6277	2.7482	0.0028	*
s(Time:%NasDur):Vowel_[e]	5.5816	7.1428	1.6152	0.1229	
s(Time:%NasDur):Vowel_[o]	5.7767	7.7293	0.4051	0.9088	
Random_Participant	188.3026	391.0000	1.3211	< 0.0001	***
Random_Item	11.3901	77.0000	0.1996	0.0097	**

Figure 5.27 **Erreur ! Nous n'avons pas trouvé la source du renvoi.** illustrates the significant %NASDUR factor for the proportions of fixations to the CVN image when listeners chose the CVN image. It is shown that the proportions of fixations to the CVN image increase overall as nasalization duration increases as well between 0N and 50N. However, between 50N and 100N nasalization duration, the proportions of fixations are similar, suggesting a more categorical pattern of recognition based on %NASDUR. In other words, when participants chose the coarticulated image and the vowel was nasalized for a short period, they tended to fixate to the CVN image less. This is unexpected, since CVN vowel are usually not nasalized, or only for

a short period of time. However, the substantial number of CVC responses in the 0N %NASDUR condition and the fact that 0N stimuli in Experiment 2b did not correspond to any image on the display may explain the lower proportions of fixations in this %NASDUR condition. It is thus possible, if one eliminates the 0N %NASDUR condition, that proportions of fixations are similar across %NASDUR conditions for L2 listeners.

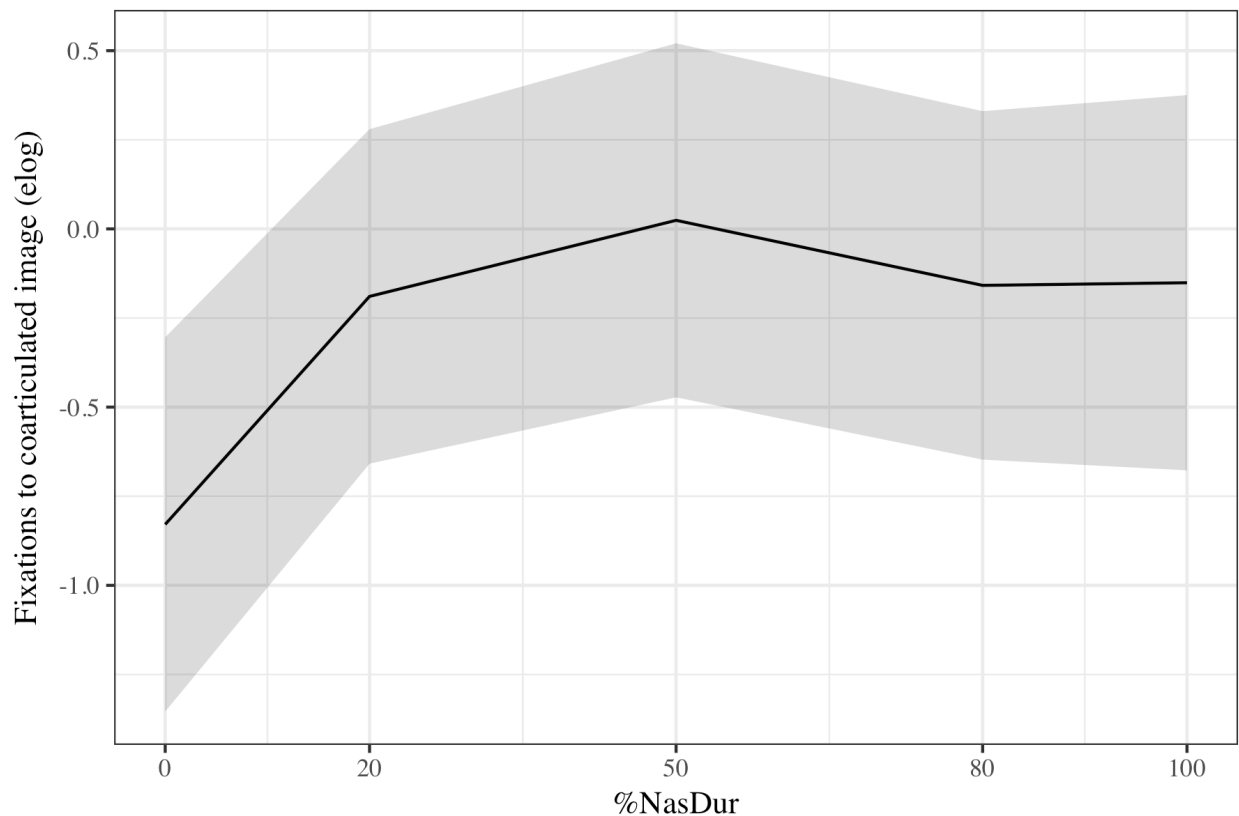


Figure 5.27. Predicted data by %NASDUR – Exp. 2b elog %fix to CVN, CVN choices

Figure 5.28 presents the predicted data for the significant %NASDURXVOWEL QUALITY through TIME interaction. The fixation patterns seem complex in the figure, suggesting that processing abilities are variable across vowel qualities and %NASDUR values. For example, the 0N and 100N %NASDUR values behave differently depending on the quality of the vowel.

However, it is difficult to draw firm conclusions regarding this complex interaction given its complexity, and the smaller number of CVN trials in Experiment 2b.

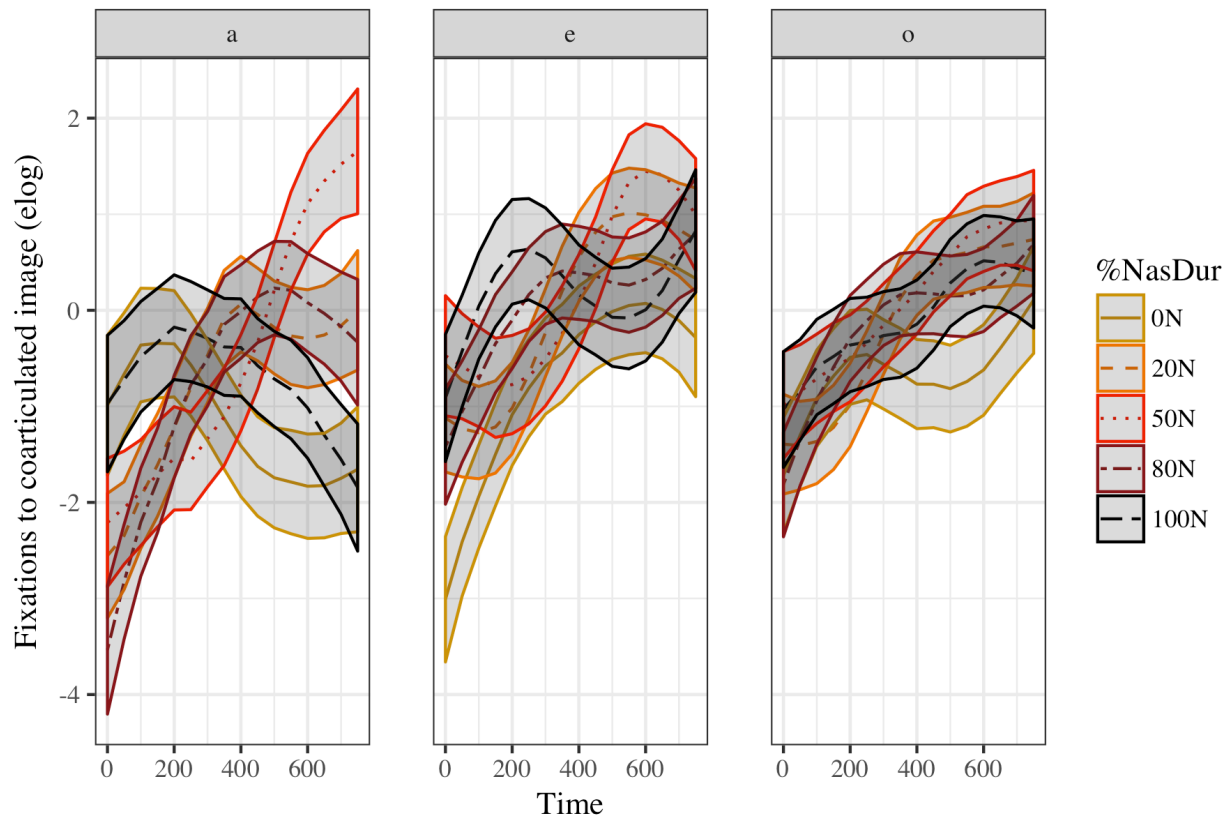


Figure 5.28. Predicted data by %NASDUR and VOWEL QUALITY through TIME – Exp. 2b elog %fix to CVN, CVN choices

5.2.3 Additional results and discussion

The results of Experiment 2b suggest that L2 listeners of CF were able to use nasalization duration as a cue to C $\tilde{V}$ and CVN words in a similar manner as L1 listeners in Experiment 1b (Section 3.2). This was shown by proportions of image choices and eye movement patterns. Similarly to Experiment 2a, listeners in Experiment 2b were able to differentiate between vowels that were not nasalized (e.g. 0N %NASDUR values) and vowels that were nasalized, even for a

short period of time (20N, 50N, 80N, and 100N %NASDUR values). They identified 0N stimuli as C $\tilde{V}$ or CVC images in a greater proportion than CVN images, since stimuli never included a nasal consonant or appendix in Experiment 2b. Also, some stimuli included formant transitions towards the oral consonant in CVC words, which explains the significant portion of CVC choices. Nasalized vowels (20N or more) were recognized as C $\tilde{V}$ between 67.74% and 79.53% of the time. The eye tracking data also supports this finding. The difference in perception of non-nasalized (0N) vs. nasalized (20N or more) vowels, when listeners chose the nasal (C $\tilde{V}$) image, suggests sensitivity to vowel nasalization in their L2. This is also consistent with Beddor et al.'s (2013) study, in which English-native listeners were able to anticipate nasal consonants based on nasalization duration information on the vowel (Beddor et al., 2013).

These results are also consistent with the hypothesis that they transferred their sensitivity to vowel nasalization duration from their native language (Major, 2008), as it has been shown that English listeners are sensitive to vowel nasalization (Beddor et al., 2013; Zamuner et al., 2016). Furthermore, the lower rate of C $\tilde{V}$ word choice when compared to L1 French listeners (Experiment 1b) suggests that their C $\tilde{V}$ representations might not be as robust as those of L1 listeners. Recall that in Experiment 1b, native listeners categorized the 20N or more %NasDur stimuli as C $\tilde{V}$ more between 81% and 95% of the time, which seems overall higher than for L2 listeners in Experiment 2b. Indeed, if listeners are not as proficient in the L2 as L1 speakers, phonetic and phonological representations may not be as accurate in the L2 and lead to overall poorer performance. However, the pattern of increase in C $\tilde{V}$ image choices between 0N and 50N %NASDUR values seems similar for L1 and L2 listeners, despite the apparent uncertainty of L2 listeners.

As mentioned above, the lower rate of C $\tilde{V}$ word choices for L2 listeners can be explained by the fact that less proficient bilinguals tend to transfer their native language's properties into

their L2 more than highly proficient bilinguals (Best & Tyler, 2007; Flege, 1995). As English does not have phonological nasal vowels in its inventory, listeners may have transferred their knowledge of a coarticulatory rule of nasalization that applies to VN sequences in their native language into L2 French, and thus interpreted long-nasalized stimuli as instances of CVN words. Furthermore, as for L1 listeners in Experiment 1b, it seems like L2 CF listeners in Experiment 2b recognized the stimuli in a categorical manner. Stimuli that were nasalized for 20% or more of their duration were recognized as C $\tilde{V}$ words well above chance, similarly to L1 listeners in Experiment 1b, and participants seemed overall less sensitive to within-category variability (e.g. between 50N and 100N %NASDUR values).

Results of Experiment 2b also suggested that L2 listeners were not able to *gradiently* perceive small differences in %NASDUR value, similarly to L1 listeners in Experiment 1b, and this can be due to two main factors. Firstly, according to the conclusions of Experiment 1a in Chapter 3, gradient perceptual patterns emerged when stimuli were ambiguous and contained conflicting phonetic cues (at least in the case of vowel nasalization). It was not the case in Experiment 2b, in which listeners were presented with auditory stimuli that corresponded to C $\tilde{V}$ words (50N or more %NASDUR), were partly consistent with the realization of a C $\tilde{V}$ word (20N %NASDUR had only slight nasalization, but no final nasal consonant), or were consistent with none of the presented word choices (0N %NASDUR). In other words, stimuli from Experiment 2b contained phonetic cues that corresponded to the canonical realization of the word or missed some phonetic cues, but never had *conflicting* phonetic cues (i.e. vowel nasalization and nasal consonant within the stimulus).

Secondly, word representations must include fine-grained phonetic details for listeners to use them during lexical access. However, recall that L2 listeners were not able to access fine-grained phonetic information in Experiment 2a when stimuli had conflicting phonetic cues

(contrary to L1 listeners in Experiment 1a). There is no reason to think that listeners from Experiment 2b were different from listeners in Experiment 2a in this regard. Consequently, it was not expected that L2 listeners in Experiment 2b would be able to access this fine-grained phonetic information.

The results of Experiment 2b thus suggest a general tendency for listeners to categorically integrate fine-grained phonetic information in their L2, especially when passed a certain threshold in nasalization duration (i.e. vowels that were nasalized for 50% or more of their duration canonically correspond to phonological nasal vowels in Canadian French). However, stimuli that would include conflicting phonetic cues or more fine-grained phonetic variation could potentially demonstrate other patterns of phonetic integration, especially if listeners are advanced in the L2, and further research will be dedicated to refining these findings. The current experiment is an exciting starting point for further investigations.

A number of additional factors were also found to have a significant effect on RTs and eye movements. For example, VOWEL QUALITY was shown to impact RTs, suggesting longer response latencies for [a-ã] words than for [ɛ-ê] words, and for [ɛ-ê] words than for [ɔ-õ] words, especially for 0N %NASDUR stimuli. This could be due to individual items, as only three word pairs were found for each vowel quality in the current study. Furthermore, phonotactically banned word-final oral [ɔ] may have influenced fixation patterns, especially in the 0N %NASDUR condition.

5.3 GENERAL DISCUSSION

Many previous studies have established that second language (L2) speech perception and word recognition are different from the native language (L1). Previous research has often focused

on identification (Escudero, 2005) and discrimination of foreign sounds in an unknown language (Best, 1995) and in languages that were acquired after the L1 (Flege, 1995). Furthermore, L2 phono-lexical recognition studies have focused on the influence of the L1 and the confusability of L2 sounds based on the native phonological system (Weber & Broersma, 2013; Weber & Cutler, 2004). However, as it has been shown that L1 listeners are able to gradiently recognize spoken words (see Experiment 1a above; McMurray et al., 2002), patterns of phonetic integration during L2 spoken word recognition have not been largely investigated. The experiments presented in this chapter aimed to document if and how duration of nasalization on vowels has an impact on word recognition in L2 Canadian French (CF), and if L2 listeners were able to gradiently integrate fine-grained phonetic details. I used a Visual World Paradigm (Huetting et al., 2011) task to assess the hypotheses that listeners would be able to use vowel nasalization as a cue to the phonological nasality contrast in L2 CF, especially since English listeners are able to use this information in their L1 (Beddor et al., 2013), and to examine whether L2 listeners gradiently or categorically perceive fine-grained phonetic information during spoken word recognition.

In Experiment 2a above, the influence of nasalization duration on word recognition was present but not as strong as L1 listeners. The fact that listeners were able to perceive differences between short-nasalized (20N) and long-nasalized (50N or more) vowels was expected based on realization of the contrast in CF (Desmeules-Trudel & Brunelle, 2018) and Beddor et al.'s (2013) results which underlined English listeners' sensitivity to variations in vowel nasalization duration. However, as the listeners' knowledge in L2 CF is not as extensive as L1 listeners (at least for some of the participants, see self-reported proficiency scores in Figure 5.16), it makes sense that they would choose the C $\tilde{V}$ items less than L1 listeners, which was also reflected in eye movements. On the other hand, they chose the CVN image more often than L1 listeners. Both these results can be explained in terms of language transfer: English only has CVN words, and no

C $\tilde{V}$ words. Therefore, if a native English listener hears a C $\tilde{V}$ or CVN auditory stimulus and categorizes it according to their native category, they should recognize it as a CVN word more since English only has CVN words.

Overall, L2 listeners seemed to display more categorical patterns during speech processing when compared to L1 listeners who displayed gradient patterns. In Experiment 1a, I showed that gradient recognition patterns emerged for native listeners when stimuli contained conflicting phonetic cues (i.e. a nasalized vowel followed by a nasal coda). This yielded to the hypothesis that L1 listeners access fine-grained phonetic information when stimuli are ambiguous, therefore that this type of phonetic detail is included in phono-lexical representations. L2 listeners did not display the same gradient pattern in Experiment 2a, therefore suggesting that L2 listeners were not able to access (as much) fine-grained phonetic information within their phono-lexical representations. This does not directly imply that fine-grained phonetic information is *not* included in L2 representations, but it seems like the type and amount of L2 exposure (Frieda & Nozawa, 2007; Kuhl & Iverson, 1995), likely among other things, could determine the amount of stored information in L2 representations. For example, if a listener has not been widely exposed to L2 French, it is expected that their phonological representations will include less (French) phonetic details, as opposed to another listener who has been thoroughly exposed. For the latter, we would expect phono-lexical representations to include more details. In sum, despite the fact that there was no strong support for the storage of fine-grained phonetic cues in L2 representations, it may depend on language exposure (and other factors of influence such as age and manner of acquisition; Sabourin et al., 2016) and further research will be necessary to thoroughly assess the influence of these factors on the quality and quantity of stored phonetic details.

Finally, a number of additional questions will need further attention in future research. For example, investigating even more fine-grained variations of the phonetic cues (i.e. vowel nasalization duration) could provide a better understanding of the level of granularity in phono-lexical representations. It is possible that the variations that were used here did not cross the listeners' perceptual boundaries. For example, if L2 listeners categorically differentiate vowels that are nasalized for more or less 40% of their duration, the %NASDUR values that were used in the current experiments were consistent with one or the other category, and gradient effects did not emerge. As in McMurray et al. (2002), gradience could emerge if stimuli were closer to the category boundary (e.g. 25N, 30N, 35N, 40N, 45N %NASDUR values). It would also be interesting to test additional listeners varying in their amount and manner of exposure to L2 French, and systematically manipulate these variables in order to have a better idea of the influence of individual factors on perception. In addition, it is important to note that behavioural responses (i.e. word choices, eye movements) do not always reflect neurological responses to stimuli (Tremblay, 2010). It would thus be very interesting to use neuroimaging techniques such as electroencephalography (EEG) or functional magnetic resonance imagery (fMRI) to uncover gradient effects of fine-grained phonetic variations in L2 spoken word recognition. The general idea that L2 listeners seem to have coarser-grained phono-lexical representations (analogously to morpho-syntax, see Clahsen & Felser, 2006), however, remains an interesting and important finding of the experiment presented above.

6 General discussion and conclusion

In this dissertation, I explored the interface between phonetics and phonology from a psycholinguistic perspective using an eye tracking paradigm, focusing on how fine-grained phonetic information influences the processing of vowels, and if the spoken word recognition system *gradiently* integrates this fine-grained phonetic variation. I showed that phonetic detail is strongly linked to spoken word recognition in both a native (L1) and a second language (L2), thus suggesting a strong link between phonetics and phono-lexical representations within listeners' minds. I also found that in some cases, detailed in the remainder of this section, phonetic information is *gradiently* integrated, and that it is *categorically* interpreted in other cases.

6.1 SUMMARY OF THE RESULTS

In Chapters 2 and 3, I presented two experiments (1a and 1b) on the processing of variations in nasalization duration on vowels during spoken word recognition with L1 listeners of Canadian French (CF). One of the main interests of this research is that CF has a phonological contrast between oral and nasal ($\tilde{V}$) vowels. Since oral vowels followed by a nasal consonant (VN) can be nasalized in CF (Desmeules-Trudel & Brunelle, 2018; Léon, 1983; Martin et al., 2001), it is important for listeners to be able to perceive differences in nasalization duration (among other cues; Carignan, 2014) to discriminate between phonological nasal ($\tilde{V}$) and coarticulated nasalized (VN) vowels. Furthermore, research on phonetic integration has mainly focused on consonantal contrasts. It was therefore necessary to investigate how listeners interpret

gradient variations in vowels to have a better idea of the way fine-grained phonetic information is processed during spoken word recognition.

Using the Visual World Paradigm (Huetting et al., 2011), I found that L1 listeners were able to use variations in nasalization duration to discriminate words that contained phonological nasal (C $\tilde{V}$) and coarticulated nasalized (CVN) vowels (Experiments 1a and 1b). I also found that they were able to *gradiently* integrate this fine-grained phonetic information (McMurray et al., 2008), especially when stimuli contained conflicting cues (Experiment 1a), thus suggesting that this type of fine-grained information is stored in phono-lexical representations (McMurray et al., 2002). A number of additional factors also turned out significant in the statistical analyses, such as experimental trial number and word frequency. This indicates that a complex set of factors influence language processing, such as the task and the choice of experimental stimuli during spoken word processing.

In Chapter 4, I presented some models of L2 speech perception (Best 1995; Best & Tyler, 2007; Flege, 1995), which are mostly based on discrimination abilities of speech sounds, and some of the processes at play during L2 spoken word recognition (Weber & Broersma, 2013; Weber & Cutler, 2004). This led to the presentation, in Chapter 5, of two additional experiments (2a and 2b) with L2 listeners of CF in order to explore patterns of phonetic integration in another population of listeners. The experiments had the same design, stimuli and procedure as the experiments with L1 listeners. The L2 listeners were all native speakers of English and had learned French as an L2 to various levels of proficiency. The interest of conducting these experiments on the integration of variations in nasalization duration with L1 English-L2 French listeners lies in the fact that English does not have phonological nasal vowels in its inventory, but does have (variable) coarticulatory nasalization of vowels when they are followed by nasal consonants (Beddor, 2009; Cohn, 1990). Consequently, in order to correctly discriminate French

phonological nasal ($\tilde{V}$) and coarticulatory nasalized (VN) vowels, L2 listeners have to learn to phonologically differentiate between long-nasalized vowels (which correspond to $\tilde{V}$ s) and short-nasalized vowels (which correspond to VNs). Learning the new *phonological-nasal-vowels* ($\tilde{V}$) category should thus enable them to avoid transfer from their L1 onto French. Based on previous literature on the perception of vowel nasalization in English (Beddor et al., 2013), it was expected that L2 listeners would be able to discriminate between long-nasalized and short-nasalized vowels in French, but the extent to which they would be able to use fine-grained variations in nasalization duration was unknown.

With L2 listeners, I found that they were able to perceive differences between long-nasalized and short nasalized vowels (Experiments 2a and 2b), similarly to L1 listeners. Artifacts of language transfer (i.e. more CVN word choices and more fixations to CVN images) were also found in both experiments, especially for lower-L2-proficiency listeners. However, even when stimuli contained conflicting phonetic cues, listeners were not able to *gradiently* integrate fine-grained variations in nasalization duration, which is opposed to L1 listeners. Two possibilities are hypothesized to explain this different pattern. Firstly, it is possible that L2 listeners are not able to access fine-grained phonetic details in their phono-lexical representations, perhaps because it is generally more demanding to complete a task in the L2 and therefore that only coarser phonetic information is used in L2 word recognition. Secondly, it is also possible that fine-grained phonetic information is not stored in L2 phono-lexical representations, and that this could depend on individual factors such as L2 proficiency, age of acquisition, or manner of acquisition (Sabourin et al., 2014). Indeed, as L2 listeners have not been exposed to the target language as much as L1 listeners, it is possible that the quality and quantity of the stored phonetic information is not as extensive in the L2 as it is in the L1.

Table 6.1 below summarizes the results of this dissertation's experiments to facilitate the comparison between experiments and populations (i.e. L1 and L2 listeners). In the table, *conflicting phonetic cues* correspond to the co-occurrence of vowel nasalization on the vowel and a nasal appendix within the stimuli, and *non-conflicting phonetic cues* correspond to the occurrence of only variations in vowel nasalization duration within stimuli.

Table 6.1

Summary of the experimental results on the patterns of phonetic integration

	Conflicting phonetic cues	Non-conflicting phonetic cues
L1 listeners	EXPERIMENT 1A: Listeners <i>gradiently</i> integrated variations in nasalization duration	EXPERIMENT 1B: Ceiling effect: Gradience did not emerge
L2 listeners	EXPERIMENT 2A: Listeners <i>did not gradiently</i> integrate variations in nasalization duration	EXPERIMENT 2B: Listeners <i>did not gradiently</i> integrate variations in nasalization duration

The remainder of this chapter will link the experimental results to the hypothesis of gradient recognition of phonetic information. This will lead to a discussion of the implications for the differences and similarities between L1 and L2 spoken word recognition and the phonetic content of phono-lexical representations. Lastly, I will suggest a number of limitations to the current work and directions for future research.

6.2 INTEGRATING L1 AND L2 SPOKEN WORD RECOGNITION

Hypotheses on the recognition of phonetic information during spoken word recognition have claimed that it can be continuous and gradient (McMurray et al., 2008). This enabled me to formulate predictions regarding listeners' behaviour on a spoken word recognition task. For

example, in an eye tracking paradigm, gradient recognition predicts that fixations to a target word or image will *gradiently* increase as the phonetic characteristics of the stimulus get closer to the canonical realization of the word. McMurray et al. (2008) found such a gradient pattern of eye movements using a VOT continuum corresponding to voiced initial consonants (e.g. as in the word *beach*; 0 ms VOT) and voiceless initial consonants (e.g. as in *peach*; 40 ms), using increments of 5 ms between each stimulus of the continuum. Their results showed that stimuli that had short VOT led to more fixations to the voiced member of the pair, and longer VOT values yielded more fixations to the voiceless member. Crucially, intermediate VOT values yielded intermediate proportions of fixations to the targets, which supports the gradient recognition hypothesis.

The results of Experiment 1a with L1 listeners also supported the predictions of the gradient recognition hypothesis (when participants interpreted the vowels as instances of the phonological nasal category), using a continuum of vowel nasalization duration. In this case, vowels contained both variations in nasalization duration and a (short) nasal coda consonant, which were conflicting (i.e. long-nasalized vowels followed by a nasal consonant are phonotactically impossible in French isolated words) when the vowel was nasalized for 50% of its duration or more. This result is especially interesting since, to my knowledge, no study has explicitly manipulated a continuum pertaining to a vowel contrast within a language to investigate the question of gradient recognition during word recognition. This suggests that gradient patterns of word recognition can also be found for vowels, therefore that they apply to a variety of contrasts, and in more than one language.

However, this pattern of gradient word recognition was not replicated with L1 listeners when the stimuli did not contain conflicting phonetic cues, and with L2 listeners in general. For L1 listeners, this can be explained by the fact that stimuli that did not contain conflicting phonetic

cues led to a ceiling effect. Indeed, if stimuli contain phonetic cues that are consistent with the canonical realization of the word, they are expected to be easy to recognize and will not have listeners access more details about the phonetics of the target. In other words, coarser phonetic information (e.g. long-nasalized vs. short-nasalized vowels, regardless of non-phonological variability) is sufficient to recognize the words.

However, the explanation is different for L2 listeners, since it has been shown that completing a task in an L2 is more difficult and demanding than in the L1 (e.g. in morphosyntax, see McDonald, 2006) due to differences in processing. It is thus possible that the L2 listeners were not able to dedicate sufficient cognitive resources to access fine-grained phonetic information in their representations of the words, especially in Experiment 2a which used the same stimuli with conflicting phonetic cues. For example, if a listener must dedicate more resources than L1 listeners to access lexical items in general, and then must associate an auditory stimulus to a representation of that lexical item (e.g. here, an image on a computer screen), it is possible that their word recognition system will “filter out” or “not pay attention to” fine-grained variations in the stimulus phonetics. Simply put, if the task in the L2 is more difficult than in the L1, the word recognition system may not consider the same amount of information in the L2 than in the L1. This could have the effect that L2 listeners display less gradient-like and more categorical patterns of recognition, which were observed in both Experiments 2a and 2b.

Another possible explanation has to do with the fact that L2 listeners often have less experience in the target language than L1 listeners (Frieda & Nozawa, 2007), which implies that their representations may not be as rich as those in the L1. Furthermore, it has been shown that the amount of exposure, manner of acquisition, L2 proficiency and age of acquisition (among other factors; Sabourin et al., 2016) all have an influence on L2 input, and thus on L2 knowledge (Barcroft & Wong, 2013; Gass, 1997). Variable individual background in L2 experience could

thus have the effect that L2 (phono-lexical) representations do not include as much phonetic details as in the L1, and therefore that these details cannot be accessed during spoken word recognition.

Although the explanations above are logically possible, they refer to two different mechanisms and are not mutually exclusive; the first one has to do with L2 processing abilities, while the second one has to do with the quality and content of L2 phono-lexical representations. Furthermore, familiar words might include more phonetic details and require less cognitive resources for processing, while unfamiliar words might be different in this regard. It is also possible that familiar words do have a good amount of phonetic details in their representations, but that completing the task in the L2 is too difficult. Task difficulty may thus play a role in recognition and phonetic details processing, especially in the L2. Further investigation will be necessary to tease them apart, for example by using a contrast that exists in the listeners' L1, and for which they would be able to use the phonetic details in their L2 (see Section 6.4 below), or by using neuroimaging techniques that are an even better reflection of low-level processing.

In conclusion, it seems like words' phonetic representations are different in L1 and L2 listeners, and that the increased cognitive load in an L2 has a major influence on the access (or storage) and processing of fine-grained phonetic details.

6.3 PHONO-LEXICAL REPRESENTATIONS INCLUDE PHONETIC DETAILS

In addition to the processing of spoken words in an L1 and an L2, the current work also examined the link between phonetics and phonology, also referred to as the phonetics-phonology interface (Hume & Johnson, 2001; Kingston, 2007; Ohala, 1990; 1996; 2010; Pierrehumbert, 2002). As mentioned above, I found that phonetic details are considered by listeners when

recognizing spoken words, suggesting that phono-lexical representations include such fine-grained phonetic details. Thus, it seems like individual abstract phonological representations, which are traditionally described as the main object of study in phonology (Chomsky & Halle, 1968), depend on fine-grained phonetic information.

This work also underlines the possibility that phonological and phonetic representations could be conflated. In the experiments above, word access performance depended on fine-grained phonetic details, it is thus reasonable to think that phonetic information is closely linked to word representations in the listeners' minds. From a psychologically-realistic point of view (Goldrick, 2011), then, it follows that phonological and lexical representations should integrate fine-grained phonetic information. Recall that other theoretical frameworks also embed phonetic information within phonological and lexical representations, such as Articulatory Phonology (Browman & Goldstein, 1986) and exemplar theories of phonology (Pierrehumbert, 2002). The current work thus adds to the body of literature on the phonetic content of phonological representations in perception and on the *dynamic* character of the phono-lexical recognition system, since time-dependent phonetic cues were considered by the listeners.

As some work in the 1980s and 1990s on phonology focused on stripping off as much phonetic information from phonological representations as possible, especially in underspecification frameworks (Archangeli, 1988; Keating, 1988; Steriade, 1995), current research has taken a new turn towards the integration of the phonetic and phonological modules of language. Notably, the emergence of frameworks such as Laboratory Phonology (Pierrehumbert, Beckman, & Ladd, 2000) has enabled researchers to find ways to reconcile the study of physical aspects of speech with the study of more abstract phonological and psychological entities. The current dissertation aimed to contribute to this endeavour, and has successfully shown that listeners processing system is able to take into account fine-grained

phonetics, thus arguing in favour of the specification of this type of information within representations.

In addition, although the TRACE model of spoken word recognition (McClelland & Elman, 1986) does not assume a specific level of granularity concerning the amount of phonetic details in the representations, gradient/dynamic phonetic integration is consistent with the architecture and mechanics of TRACE. For example, as has already been shown by McMurray et al. (2009) with VOT variations, the current results suggest that listeners consider the phonetic information (i.e. earlier or later onset of nasalization within vowels) as the speech signal unfolds, in both an L1 and an L2. Early onset of nasalization further activates C $\tilde{V}$ words, and late onset of nasalization further activates CVN words. This shows that representations not only include fine-grained phonetic information, but that they may also include a *timing* component because listeners were sensitive to the moment at which the phonetic cue was presented.

6.4 LIMITATIONS AND FURTHER RESEARCH

Finally, a number of limitations are inevitable when investigating complex questions such as the phonetics-phonology interface, the integration of fine-grained phonetic information, and L2 phono-lexical processing. Most of these limitations lead to new research questions for further investigations. In the current section, I will describe a number of limitations to the experiments that I conducted, as well as future directions for research on phonetic detail processing and phono-lexical representations.

Firstly, recall that the experiments that I presented above are among the first to explore patterns of phonetic integration using a vowel contrast. Consequently, even though the conclusions that I was able to draw fit within the literature on gradient perception that was based

on consonant contrasts, it will be important in the future to explore a greater number of vocalic properties to have a more thorough idea of how generalizable the patterns are that were reported here. For example, creating a continuum in formant values would enable me to explore height (e.g. from [e] to [ɛ]), anteriority (e.g. from [y] to [u]) or rounding (e.g. from [i] to [y]) contrasts in vowels. Using an eye tracking paradigm that is similar to the one I used and which enables us to dynamically track the speech comprehension, I would also be able to further document the gradient vs. categorical patterns of vowel perception (Fry, Abramson, Eimas, & Liberman, 1962), as opposed to consonants which were shown to be perceived more categorically (Liberman et al., 1957).

Furthermore, I have underlined two possibilities to explain the result patterns of L2 listeners: their inability to gradiently recognize variable spoken words can be due to increased cognitive load for L2 processing which prevents them from accessing and using fine-grained phonetic information, or due to the absence of such specified phonetic details in their representations. To test which explanation better reflects L2 representations, it would be informative to investigate L2 listeners using a contrast that already exists in their L1 but with L2 lexical items, e.g. L1 English listeners who must recognize minimal pairs of words that contain [ɛ]-[a] in L2 French. If the *increased-processing-load* explanation is true, L2 listeners should display categorical word recognition patterns for [ɛ]-[a], since cognitive demands in the L2 are greater and should have the same categorical effect even if the phonetic contrast is mastered in the L2. However, if the *phonetic-content* explanation is true, they should display gradient patterns because their representations already include phonetic details about the difference between [ɛ] and [a] (due to language transfer). With early L2 bilinguals or listeners who are highly proficient in their L2, it is expected that they would respond more similarly to L1 listeners (i.e. more gradient recognition patterns), showing that their representations include fine-grained phonetic

details. For beginners and late bilinguals, however, it is expected that they would display more categorical effects, perhaps because the task would be more difficult for them to complete. In other words, I would expect that more proficient listeners would have more phonetic details in their representations and would display gradient patterns, while less proficient listeners would display categorical patterns due to increased processing load.

Finally, as the theoretical claims that I have made in this dissertation are based on the nasality contrast in French with L1 and L2 listeners, it will be essential to investigate a greater variety of languages (and language combinations) in the future to verify if and how listeners with various language backgrounds process fine-grained phonetic details. Despite the robust evidence in support of the inclusion of phonetics within phono-lexical representations, a number of additional variables have also been found to have a significant effect, such as other properties of the contrast and listeners' language backgrounds (e.g. province of origin, proficiency in the L2, early vs. late exposure to the L2). Investigations of other languages will thus be crucial to our understanding of the linguistic processing system in general.

In sum, I have shown that native listeners of Canadian French were able to gradiently recognize spoken words that varied in nasalization duration on the vowel, using eye tracking and stimuli that contained conflicting phonetic cues. These conflicting cues made the word recognition task harder, which probably pushed the listeners to access more fine-grained phonetic information, suggesting that it is specified in the phono-lexical representations. When stimuli did not have conflicting cues, a ceiling effect was observed, and gradience in word recognition did not emerge. I also found that second language listeners of Canadian French (native English) were not able to gradiently recognize spoken words, but were still able to (coarsely) differentiate words that contained phonological nasal vowels and coarticulatory nasalized vowels. This can be due to their inability to access such fine-grained phonetic information in their L2 because of task

difficulty, or to the lack of specification of fine-grained phonetic information; more research will be necessary to tease those hypotheses apart. Language transfer was also observed, especially in lower-proficiency listeners. These experiments provided support in favour of the gradient recognition of phonetic information (McMurray et al., 2008), thus suggesting a strong link between phonetics and phonology.

References

- Alamri, A. A. (2017). *Phonological, Semantic, and Root Activation in Spoken Word Recognition in Arabic : Evidence from Eye Movements* (Unpublished doctoral dissertation). University of Ottawa, Ottawa, ON.
- Allen, J. S., & Miller, J. L. (1999). Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. *The Journal of the Acoustical Society of America*, 106(4), 2031–2039.
- Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38, 419–439.
- Andrews, S., Miller, B., & Rayner, K. (2004). Eye movements and morphological segmentation of compound words: There is a mouse in mousetrap. *European Journal of Cognitive Psychology*, 16(1–2), 285–311.
- Archangeli, D. (1988). Aspects of underspecification theory. *Phonology*, 5, 183–207.
- Archila-Suerte, P., Zevin, J., Bunta, F., & Hernandez, A. E. (2012). Age of acquisition and proficiency in a second language independently influence the perception of non-native speech. *Bilingualism*, 15(1), 190–201.
- Arnaud, V. (2003). French nasal vowels spoken in Haut-Jura: a quantitative acoustic study. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 929–932). Barcelona, Spain.

- Baker, A., Archangeli, D., & Mielke, J. (2011). Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change*, 23(3), 347–374.
- Barcroft, J., & Wong, W. (2013). Input, input processing and focus on form. In J. Herschensohn & M. Young-Scholten (Eds.), *The Cambridge Handbook of Second Language Acquisition* (pp. 627–647). Cambridge, UK: Cambridge University Press.
- Barr, D. J. (2008). Analyzing “visual world” eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59(4), 457–474.
- Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, 85(4), 785–821.
- Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *Journal of the Acoustical Society of America*, 133(4), 2350–2366.
- Best, C. T. (1995). A direct realistic view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 171–204). Timonium, MD: York Press.
- Best, C. T., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech*, 46(2–3), 183–216.
- Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener’s native phonological system. *The Journal of the Acoustical Society of America*, 109(2), 775–794.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language*

- Experience in Second Language Speech Learning: In Honor of James Emil Flege* (pp. 13–34). Amsterdam, NL: John Benjamins.
- Bialystok, E. (2015). The impact of bilingualism on cognition. In R. Scott & S. Kosslyn (Eds.), *Emerging Trends in the Social and Behavioral Sciences* (pp. 1–12). Hoboken, NJ: John Wiley and Sons.
- Bohn, O., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, 11, 303–328.
- Brannen, K. (2002). The role of perception in differential substitution. *Canadian Journal of Linguistics*, 47(2), 1–46.
- Brien, C. (2013). *Neurophysiological evidence of a second language influencing lexical ambiguity resolution in the first language* (Unpublished doctoral dissertation). University of Ottawa, Ottawa, ON.
- Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology*, 3, 219–252.
- Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011). The assimilation of L2 Australian English vowels to L1 Japanese vowel categories: Vocabulary size matters. *Applied Psycholinguistics*, 32, 51–67.
- Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE*, 5(6).
- Carignan, C. (2014). An acoustic and articulatory examination of the “oral” in “nasal”: The oral articulations of French nasal vowels are not arbitrary. *Journal of Phonetics*, 46(1), 23–33.

- Carignan, C. M. (2013). *When nasal is more than nasal: the oral articulation of nasal vowels in two dialects of French* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign, Urbana-Champaign, IL.
- Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. New York, NY: Harper & Row.
- Clahsen, H., & Felser, C. (2006). How native-like is non-native language processing? *Trends in Cognitive Sciences*, 10(12), 564–570.
- Cohn, A. C. (1990). *Phonetic and phonological rules of nasalization* (Unpublished doctoral dissertation). University of California Los Angeles, Los Angeles, CA.
- Cohn, A. C. (2006). Is there gradient phonology? In G. Fanselow, M. Féry, M. Schlesewsky, & R. Vogel (Eds.), *Gradience in Grammar: Generative Perspectives* (pp. 143–204). Oxford, UK: Oxford University Press.
- Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, 6, 84-107.
- Coquillon, A., & Turcsan, G. (2012). An overview of the phonological and phonetic properties of Southern French: Data from two Marseille surveys. In R. Gess, C. Lyche, & T. Meisenburg (Eds.), *Phonological Variation in French: Illustrations from Three Continents* (pp. 105–127). Amsterdam, NL: John Benjamins.
- Côté, M.-H. (2012). Laurentian French (Québec): Extra vowels, missing schwas, and surprising liaison consonants. In R. Gess, C. Lyche, & T. Meisenburg (Eds.), *Phonological Variation*

- in French: Illustrations from Three Continents* (pp. 235–274). Amsterdam, NL: John Benjamins.
- Cuetos, F., Glez-Nosti, M., Barbón, A., & Brysbaert, M. (2011). SUBTLEX-ESP: Spanish word frequencies based on film subtitles. *Psicológica*, 32, 133–143.
- Cutler, A. (1987). Components of prosodic effects in speech recognition. In *Proceedings of the 11th International Congress of Phonetic Sciences* (pp. 84–87). Tallin, Estonia.
- Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001a). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, 42(4), 317–367.
- Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001b). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, 16(5–6), 507–534.
- Dang, J., & Honda, K. (1996). Acoustic characteristics of the human paranasal sinuses derived from transmission characteristic measurement and morphological observation. *The Journal of the Acoustical Society of America*, 100(5), 3374–3383.
- Dang, J., Honda, K., & Suzuki, H. (1994). Morphological and acoustical analysis of the nasal and the paranasal cavities. *The Journal of the Acoustical Society of America*, 96(4), 2088–2100.
- Delattre, P. (1965). La nasalité vocalique en français et en anglais. *The French Review*, 39(1), 92–109.
- Delvaux, V. (2006). Production des voyelles nasales en français québécois. In *Actes Des 26^{es} Journées d'Études Sur La Parole* (pp. 383–386). Dinard, France.

- Desmeules-Trudel, F. (2013). *Perception des voyelles nasales du français québécois : Aspects acoustiques et perceptifs* (Unpublished masters memoir). Université Laval, Québec City, QC.
- Desmeules-Trudel, F. (2015a). Propriétés aérodynamiques des voyelles nasales et potentiellement nasalisées en français québécois. In S. Vinerte (Ed.), *Proceedings of the 2015 Annual Conference of the Canadian Linguistics Association*. Ottawa, ON.
- Desmeules-Trudel, F. (2015b). The aerodynamics of vowel nasality and nasalization in Brazilian Portuguese. In The Scottish Consortium for ICPhS (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences* (paper number 769). Glasgow, Scotland.
- Desmeules-Trudel, F., & Brunelle, M. (2018). Phonotactic restrictions condition the realization of vowel nasality and nasal coarticulation: Duration and airflow measurements in Québécois French and Brazilian Portuguese. Manuscript submitted to *Journal of Phonetics*.
- Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5(3). 175–197.
- Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychonomic Bulletin & Review*, 15(4), 850–855.
- Escudero, P. (2005). *Linguistic Perception and Second Language Acquisition: Explaining the attainment of Optimal Phonological Categorization* (Doctoral dissertation). *LOT Dissertation Series 113*. Utrecht University, Utrecht, NL.

- Escudero, P., & Chládková, K. (2010). Spanish listeners' perception of American and Southern British English vowels. *The Journal of the Acoustical Society of America*, 128(5), EL254-EL260.
- Escudero, P., Sisinni, B., & Grimaldi, M. (2014). The effect of vowel inventory and acoustic properties in Salento Italian learners of Southern British English vowels. *The Journal of the Acoustical Society of America*, 135(3), 1577–1584.
- Escudero, P., & Vasiliev, P. (2011). Cross-language acoustic similarity predicts perceptual assimilation of Canadian English and Canadian French vowels. *The Journal of the Acoustical Society of America*, 130(5), EL277-EL283.
- Escudero, P., & Williams, D. (2011). Perceptual assimilation of Dutch vowels by Peruvian Spanish listeners. *The Journal of the Acoustical Society of America*, 129(1), EL1-EL7.
- Farnetani, E., & Recasens, D. (2010). Coarticulation and connected speech processes. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (2nd ed., pp. 316–352). Malden, MA: Wiley-Blackwell.
- Feng, G., & Castelli, E. (1996). Some acoustic features of nasal and nasalized vowels: A target for vowel nasalization. *The Journal of the Acoustical Society of America*, 99(6), 3694–3706.
- Fischer, B. (1992). Saccadic reaction time: Implications for reading, dyslexia, and visual cognition. In K. Rayner (Ed.), *Eye movements and visual cognition* (pp. 31–45). New York, NY: Springer.
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research I* (pp. 233–277). Timonium, MD: York Press.

- Flege, J. E., & Mackay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, 26, 1–34.
- Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, 12(6), 627–635.
- Frieda, E. M., & Nozawa, T. (2007). You are what you eat phonetically: The effect of linguistic experience on the perception of foreign vowels. In O.-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege* (pp. 79–96). Amsterdam, NL: John Benjamins.
- Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, 5(4), 171–189.
- Fujimura, O. (1962). Analysis of nasal consonants. *The Journal of the Acoustical Society of America*, 34(12), 1865–1875.
- Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6(1), 110–125.
- Garrett, A., & Johnson, K. (2013). Phonetic bias in sound change. In A. C. Yu (Ed.), *Origins of Sound Change: Approaches to Phonologization* (pp. 51-97). Oxford, UK: Oxford University Press.
- Gass, S. M. (1997). *Input, Interaction, and the Second Language Learner*. Hillsdale, NJ: Erlbaum.
- Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, 66(3), 363–376.

- Gilichinskaya, Y. D., & Strange, W. (2010). Perceptual assimilation of American English vowels by inexperienced Russian listeners. *The Journal of the Acoustical Society of America*, 128(2), EL80-EL85.
- Goldrick, M. (2011). Using psychological realism to advance phonological theory. In A. C. L. Yu (Ed.), *The Handbook of Phonological Theory* (2nd ed., pp. 632–660). Malden, MA: Wiley-Blackwell.
- Goldstone, R. L., & Hendrickson, A. T. (2009). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(1), 69–78.
- Gow, D. W. (2003). Feature parsing: feature cue mapping in spoken word recognition. *Perception & Psychophysics*, 65(4), 575–590.
- Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America*, 107(5), 2711–2724.
- Hall, D. C., & Hall, K. C. (2016). Marginal contrasts and the Contrastivist Hypothesis. *Glossa: A Journal of General Linguistics*, 1(1), 1–23.
- Hall, K. C. (2013). A typology of intermediate phonological relationships. *Linguistic Review*, 30(2), 215–275.
- Harnad, S. (2003). Categorical perception. In *Encyclopedia of Cognitive Science* (pp. 448–452). Malden, MA: Wiley-Blackwell.
- Huetig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137(2), 151–171.

- Hume, E., & Johnson, K. (2001). *The Role of Speech Perception in Phonology*. San Diego, CA: Academic Press.
- Keating, P. A. (1988). Underspecification in phonetics. *Phonology*, 5, 275–292.
- Kessinger, R. H., & Blumstein, S. E. (1998). Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics*, 26(2), 117–128.
- Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, 42(3), 643–650.
- Kingston, J. (2007). The phonetics phonology interface. In P. V. De Lacy (Ed.), *The Cambridge Handbook of Phonology* (pp. 401–434). New York, NY: Cambridge University Press.
- Koirala, C. (2015). The word frequency effect on second language vocabulary learning. In F. Helm, L. Bradley, M. Guarda, & S. Thouësny (Eds.), *Critical CALL - Proceedings of the 2015 EUROCALL Conference* (pp. 318–323). Dublin, Ireland: Research-publishing.net.
- Krakow, R. A. (1999). Physiological organization of syllables: A review. *Journal of Phonetics*, 27(1), 23–54.
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363, 979–1000.
- Kuhl, P. K. (1992). Psychoacoustics and speech perception: Internal standards, perceptual anchors, and prototypes. In L. A. Werner & E. W. Rubel (Eds.), *Developmental Psychoacoustics* (pp. 293–332). Washington, DC: American Psychological Association.

- Kuhl, P. K. (1987). The special-mechanisms debate in speech research: Categorization tests on animals and infants. In S. Harnad (Ed.), *Categorical Perception: The Groundwork of Cognition* (pp. 355–386). New York, NY: Cambridge University Press.
- Kuhl, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology*, 4(6), 812–822.
- Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the “perceptual magnet effect.” In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 121–154). Timonium, MD: York Press.
- Kurowski, K. M., & Blumstein, S. E. (1993). Acoustic properties for the perception of nasal consonants. In M. K. Huffman & R. A. Krakow (Eds.), *Nasals, Nasalization, and the Velum* (pp. 197–222). London, UK: Academic Press.
- Ladd, D. R. (2011). Phonetics in phonology. In J. Goldsmith, J. Riggle, & A. C. Yu (Eds.), *The Handbook of Phonological Theory* (2nd ed., pp. 348–373). Malden, MA: Wiley-Blackwell.
- Lahiri, A., & Marslen-Wilson, W. (1990). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, 38, 245–294.
- Léon, P. R. (1983). Les voyelles nasales et leurs réalisations dans les parlers français du Canada. *Langue Française*, 60, 48-64.
- Levy, E. S. (2009). On the assimilation-discrimination relationship in American English adults’ French vowel learning. *The Journal of the Acoustical Society of America*, 126(5), 2670–2682.
- Levy, E. S., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics*, 36(1), 141–157.

- Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54(5), 358–368.
- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Haskins Laboratory Status Report on Speech Research*, 82–83, 63–93.
- Luce, P. A. (1986). *Response Times*. New York, NY: Oxford University Press.
- Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics*, 62(3), 615–625.
- Maeda, S. (1993). Acoustics of vowel nasalization and articulatory shift in French nasal vowels. In M. K. Huffman & R. A. Krakow (Eds.), *Nasals, Nasalization, and the Velum* (pp. 147–167). London, UK: Academic Press.
- Magloughlin, L. (2018). */tɹ/ and /dɹ/ in North American English: Phonologization of a Coarticulatory Effect* (Unpublished doctoral dissertation). University of Ottawa, Ottawa, ON.
- Major, R. C. (2008). Transfer in second language phonology: A review. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 63–94). Amsterdam, NL: John Benjamins.
- Martin, P. (2002). Le système vocalique du français du Québec: De l’acoustique à la phonologie. *La Linguistique*, 38(2), 71–88.
- Martin, P., Beaudoin-Bégin, A.-M., Goulet, M.-J., & Roy, J.-P. (2001). Les voyelles nasales en français du Québec. *La Linguistique*, 37(2), 49–70.

- Max, L., & Onghena, P. (1999). Some Issues in the Statistical Analysis of Completely Randomized and Repeated Measures Designs for Speech, Language, and Hearing Research. *Journal of Speech Language and Hearing Research*, 42(2), 261–270.
- McClelland, J., & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
- McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, 55(3), 381–401.
- McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin and Review*, 15(6), 1064–1071.
- McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, 86(2), B33–B42.
- McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2009). Within-category VOT affects recovery from “lexical” garden-paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language*, 60(1), 65–91.
- McQueen, J. M., & Cutler, A. (2010). Cognitive processes in speech perception. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (2nd ed., pp. 489–520). Malden, MA: Wiley-Blackwell.
- McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon, *Cognitive Science*, 30, 1113–1126.

- Mielke, J. (2003). The interplay of speech perception and phonology: Experimental evidence from Turkish. *Phonetica*, 60(3), 347-374.
- Mielke, J. (2013). Ultrasound and corpus study of a change from below: Vowel rhoticity in Canadian French. *University of Pennsylvania Working Papers in Linguistics*, 19(2), 141–150.
- New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics*, 28(4), 661–677.
- Newman, R. S., Sawusch, J. R., & Luce, P. A. (1997). Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance*, 23(3), 873-889.
- Norris, D. (1994). Shortlist: A connectionist model of continuous model of speech recognition. *Cognition*, 52(3), 189–234.
- Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech: Feedback is never necessary. *Behavioral and Brain Sciences*, 23, 299–370.
- Ohala, J. J. (1990). There is no interface between phonetics and phonology: A personal view. *Journal of Phonetics*, 18, 153–171.
- Ohala, J. J. (2005). Phonetics and historical phonology. In B. D. Joseph, & R. D. Janda (Eds.), *The Handbook of Historical Linguistics* (2nd ed., pp. 669-686). Malden, MA: Wiley-Blackwell.
- Ohala, J. J. (2010). The relation between phonetics and phonology. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (2nd ed., pp. 653–677). Malden, MA: Wiley-Blackwell.

- Ohala, J. J. (1996). Speech perception is hearing sounds, not tongues. *The Journal of the Acoustical Society of America*, 99(3), 1718–1725.
- Ohala, J. J., & Ohala, M. (1993). The phonetics of nasal phonology: Theorems and data. In M. K. Huffman & R. A. Krakow (Eds.), *Nasals, Nasalization, and the Velum* (pp. 225–249). London, UK: Academic Press.
- Paquette-Smith, M., Fecher, N., & Johnson, E. K. (2016). Two-year-olds’ sensitivity to subphonemic mismatch during online spoken word recognition. *Attention, Perception, and Psychophysics*, 78(8), 2329–2340.
- Pierrehumbert, J. B. (2003). Phonetic Diversity, Statistical Learning, and Acquisition of Phonology. *Language and Speech*, 46(2–3), 115–154.
- Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 101–139). Berlin: Mouton De Gruyter.
- Pierrehumbert, J. B., Beckman, M., & Ladd, D. R. (2000). Conceptual foundations of phonology as a laboratory science. In N. Burton-Roberts, P. Carr, & G. Docherty (Eds.), *Phonological Knowledge: Conceptual and Empirical Issues* (pp. 273–304). Oxford, UK: Oxford University Press.
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13, 253–260.
- Porretta, V., Tucker, B. V., & Järvikivi, J. (2016). The influence of gradient foreign accentedness and listener experience on word recognition. *Journal of Phonetics*, 58, 1–21.
- Porretta, V., Kyröläinen, A., van Rij, J., & Järvikivi, J. (2018). Visual World Paradigm data: From preprocessing to nonlinear time-course analysis. In I. Czarnowski, R. J. Howlett, & L. C.

- Jain (Eds.), *Intelligent Decision Technologies 2017: Proceedings of the 9th KES International Conference on Intelligent Decision Technologies – Part II* (pp. 268-277). Springer International Publishing.
- R Core Team. (2017). The R Project for Statistical Computing [computer software]. Retrieved from <https://www.r-project.org>
- Rubin, P., Turvey, M. T., & Van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perception & Psychophysics*, 19(5), 394–398.
- Sabourin, L., Leclerc, J.-C., Burkholder, M., & Brien, C. (2014). Bilingual lexical organization: Is there a sensitive period? In L. Teddman (Ed.), *Proceedings of the 2014 Annual Conference of the Canadian Linguistics Association*. St. Catharines, ON
- Sabourin, L., Leclerc, J.-C., Lapierre, M., Burkholder, M., & Brien, C. (2016). The language background questionnaires in L2 research: Teasing apart the variables. In L. Hracs (Ed.), *Proceedings of the 2016 Annual Conference of the Canadian Linguistics Association*. Calgary, AB.
- Sabourin, L., & Vinerte, S. (2018). Cognitive control among immersed bilinguals: Considering differences in linguistic and nonlinguistic processing. *Bilingualism: Language and Cognition, First View articles*, 16 pages.
- Salverda, A. P., & Altmann, G. T. M. (2011). Attentional capture of objects referred to by spoken language. *Journal of Experimental Psychology: Human Perception and Performance*, 37(4), 1122-1133.

- Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, 71(1), 145–163.
- Salverda, A. P., Brown, M., & Tanenhaus, M. K. (2011). A goal-based perspective on eye movements in visual world studies. *Acta Psychologica*, 137(2), 172-180.
- Salverda, A. P., & Tanenhaus, M. K. (2017). The Visual World Paradigm. In A. M. B. de Groot, & P. Hagoort (Eds.), *Research Methods in Psycholinguistics and the Neurobiology of Language: A Practical Guide* (pp. 89-110). Oxford, UK: Wiley-Blackwell.
- Schouten, B., Gerrits, E., & Van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, 41, 71–80.
- Segui, J., & Frauenfelder, U. H. (1986). The effect of lexical constraints upon speech perception. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive capabilities: mechanisms and performances. Symposium in memoriam Hermann Ebbinghaus 1885* (pp. 795–808). Amsterdam, NL: Elsevier Science Publishers.
- Sherkina, M. (2003). The cognate facilitation effect in bilingual speech processing. *Toronto Working Papers in Linguistics*, 21, 135–151.
- Solé, M.-J. (1995). Spatio-temporal patterns of velopharyngeal action in phonetic and phonological nasalization. *Language and Speech*, 38(1), 1–23.
- Steriade, D. (1995). Underspecification and markedness. In J. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 114–175). Malden, MA: Wiley-Blackwell.
- Stevenson, S., & Zamuner, T. (2017). Gradient phonological relationships: Evidence from vowels in French. *Glossa: A Journal of General Linguistics*, 2(1), 1–22.

- Székely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., Lu, C. C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotchewa, T., Orozco-Figueroa, A., Tzeng, A., Tzeng, O., Arévalo, A., Vargha, A., Butler, A. C., Buffington, R., & Bates, E. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language*, 51(2), 247–250.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634.
- Thordardottir, E. (2015). The relationship between bilingual exposure and morphosyntactic development. *International Journal of Speech-Language Pathology*, 15(4), 426–445.
- Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science*, 21(10), 1532–1540.
- Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research: “Clozing” the gap. *Studies in Second Language Acquisition*, 33(3), 339–372.
- Tremblay, A., & Newman, A. J. (2015). Modeling nonlinear relationships in ERP data using mixed-effects regression with R examples. *Psychophysiology*, 52(1), 124–139.
- Tremblay, M.-C. (2010). *Comparing the Perceptual Abilities of Monolinguals, Bilinguals and Monolinguals: A Combined Behavioral and Event-Related Potential Experiment* (Unpublished doctoral dissertation). University of Ottawa, Ottawa, ON.

- Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica*, 71(1), 4–21.
- Vajta, K. (2012). Autant en emporte le vin, ou : De l'importance des voyelles nasales. *Moderna Språk*, 1, 145–156.
- van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, 67(6), 1176–1190.
- van Rij, J., Hollebrandse, B., & Hendriks, P. (2016). Children's eye gaze reveals their use of discourse context in object pronoun resolution. In A. Holler & K. Suckow (Eds.), *Empirical perspectives on anaphora resolution* (pp. 267–293). Berlin, Germany: De Gruyter.
- van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2016). itsadug: Interpreting time series and autocorrelated data using GAMMs [R package].
- Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, 167, 392–393.
- Weber, A., & Broersma, M. (2013). Spoken word recognition in second language acquisition. In *The Encyclopedia of Applied Linguistics*. Malden, MA: Wiley-Blackwell.
- Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50(1), 1–25.
- Werker, J. F., & Curtin, S. (2005). PRIMIR: A Developmental Framework of Infant Speech Processing. *Language Learning and Development*, 1(2), 197–234.
- Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America*, 75(6), 1866–1878.

- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer.
- Wode, H. (1994). Nature, nurture, and age in language acquisition: The case of speech perception. *Studies in Second Language Acquisition*, 16, 325–345.
- Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R* (2nd ed.). Boca Raton, FL: Chapman and Hall/CRC.
- Yamada, R. A. (1995). Age and acquisition of second language speech sounds: Perception of American English /r/ and /l/ by native speakers of Japanese. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 305–320). Timonium, MD: York Press.
- Zamuner, T. S., Moore, C., & Desmeules-Trudel, F. (2016). Toddlers' sensitivity to within-word coarticulation during spoken word recognition: Developmental differences in lexical competition. *Journal of Experimental Child Psychology*, 152, 136–148.

Appendices

APPENDIX 1

Language background questionnaire (French and English translation)

Code d'identification : _____

Questionnaire d'informations relatives au participant

Nom et prénom (facultatif)				Adresse courriel (facultatif)		
Âge	Sexe			Date		
	M ☐	F ☐	Autre _____			
Ville d'origine			Durée de résidence			
Ville de résidence			Durée de résidence			
Langue maternelle			Langue maternelle des parents/tuteurs			
			Mère/Tuteur 1 : Père/Tuteur 2 :			
Langue(s) seconde(s) parlée(s) (compréhension orale)		Niveau				
		1 ☹	2	3 ☹	4	5 ☺
		☐	☐	☐	☐	☐
		☐	☐	☐	☐	☐
		☐	☐	☐	☐	☐
Langue(s) seconde(s) parlée(s) (expression orale)		Niveau				
		1 ☹	2	3 ☹	4	5 ☺
		☐	☐	☐	☐	☐
		☐	☐	☐	☐	☐
		☐	☐	☐	☐	☐
Niveau de scolarité			Occupation et domaine			

Entre les âges de 0 et 5 ans

Langue	% usage à la garderie		% usage avec la famille		% usage avec les amis	
	Comp.	Expr.	Comp.	Expr.	Comp.	Expr.

Entre les âges de 5 et 13 ans (école primaire et *middle school*)

Langue	% usage à l'école		% usage avec la famille		% usage avec les amis	
	Comp.	Expr.	Comp.	Expr.	Comp.	Expr.

Entre les âges de 13 et 18 ans (école secondaire)

Langue	% usage à l'école		% usage au travail		% usage avec la famille		% usage avec les amis		% usage avec partenaire	
	Comp.	Expr.	Comp.	Expr.	Comp.	Expr.	Comp.	Expr.	Comp.	Expr.

À l'université (et au cégep)

Langue	% usage à l'école		% usage au travail		% usage avec la famille		% usage avec les amis		% usage avec partenaire	
	Comp.	Expr.	Comp.	Expr.	Comp.	Expr.	Comp.	Expr.	Comp.	Expr.

Si vous avez complété votre cheminement pré-universitaire en dehors du Québec

À l'école primaire, j'ai complété (encerclez) :

cours réguliers de français / programme d'immersion française / école de langue française

À l'école secondaire, j'ai complété (encerclez) :

cours réguliers de français / programme d'immersion française / école de langue française

Identification code: _____

Participant's information questionnaire

Surname and name (optional)				Email (optional)		
Age	Gender			Date		
	M ☐	F ☐	Other _____			
City of origin			Length of residence			
City of residence			Length of residence			
Native language			Parents'/Tutors' native language			
			Mother/Tutor 1:		Father/Tutor 2:	
Spoken second language(s) (Spoken comprehension)		Level				
		1 ☹ ☐	2 ☐	3 ☹ ☐	4 ☐	5 ☺ ☐
		☐	☐	☐	☐	☐
		☐	☐	☐	☐	☐
		☐	☐	☐	☐	☐
Spoken second language(s) (Spoken production)		Level				
		1 ☹ ☐	2 ☐	3 ☹ ☐	4 ☐	5 ☺ ☐
		☐	☐	☐	☐	☐
		☐	☐	☐	☐	☐
		☐	☐	☐	☐	☐
Level of education			Occupation and field			

Between ages 0 and 5

Language	% use in kindergarten		% use with family		% use with friends	
	Comp.	Prod.	Comp.	Prod.	Comp.	Prod.

Between ages 5 and 13 (elementary and middle school)

Language	% use in school		% use with family		% use with friends	
	Comp.	Prod.	Comp.	Prod.	Comp.	Prod.

Between ages 13 and 18 (high school)

Language	% use at school		% use at work		% use with family		% use with friends		% use with partner	
	Comp.	Prod.	Comp.	Prod.	Comp.	Prod.	Comp.	Prod.	Comp.	Prod.

At university (and cégep)

Language	% use at school		% use at work		% use with family		% use with friends		% use with partner	
	Comp.	Prod.	Comp.	Prod.	Comp.	Prod.	Comp.	Prod.	Comp.	Prod.

If you were educated outside Quebec before university

In elementary school, I completed (circle):

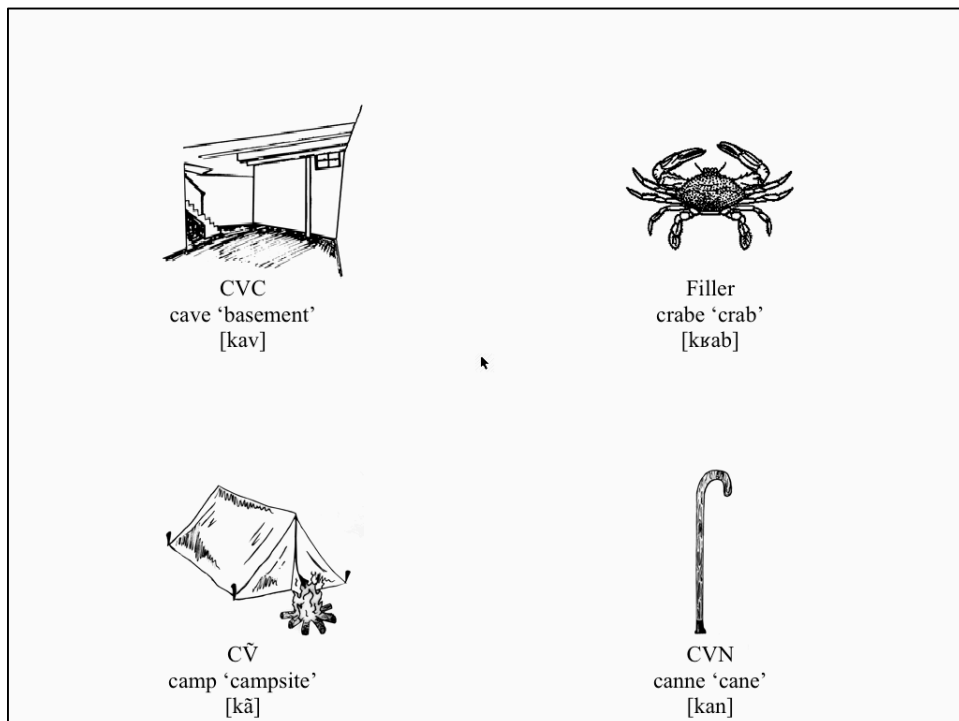
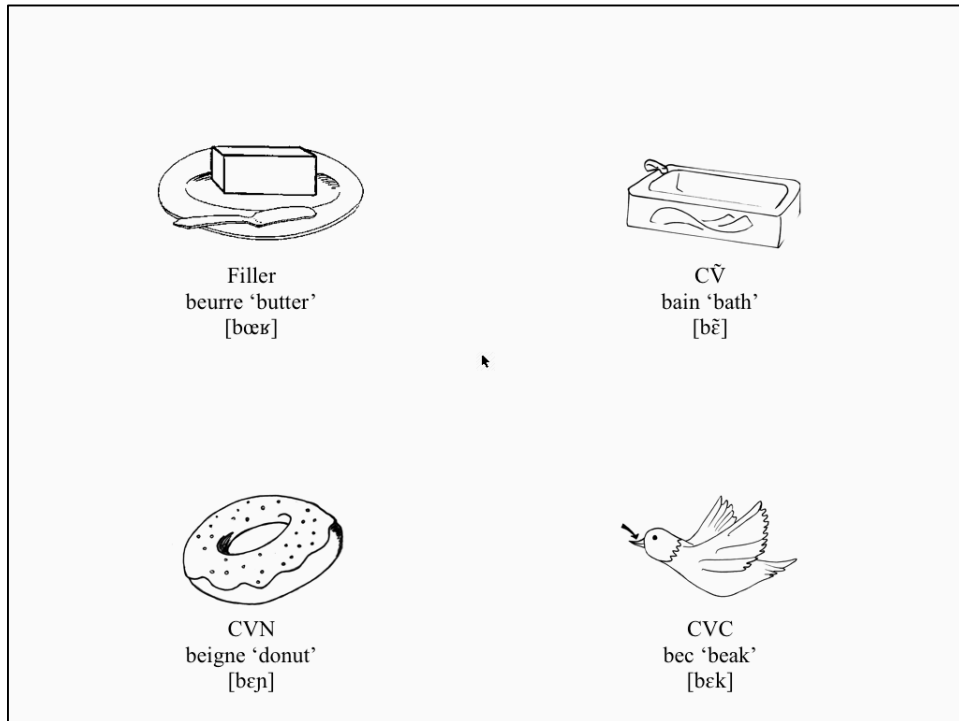
Core (regular) French courses / French immersion program / French-language school

In high school, I completed (circle):

Core (regular) French courses / French immersion program / French-language school

APPENDIX 2

Experimental displays (in alphabetical order). Note that orthographic words were not presented on the screen with the images, but are indicated here for illustration purposes.





CVN
dame 'lady'
[dam]



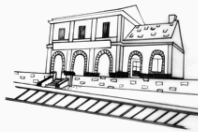
CVC
datte 'date (fruit)'
[dat]



Filler
nid 'nest'
[ni]



C $\tilde{V}$
dent 'tooth'
[dâ]



CVC
gare 'bus/train station'
[gaʁ]



Filler
gaufre 'waffle'
[gofʁ]



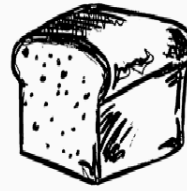
C $\tilde{V}$
gant 'glove'
[gâ]



CVN
gamme 'musical scale'
[gam]



Filler
pluie 'rain'
[plɥi]



CV̆
pain 'bread'
[pɛ̃]



CVN
peigne 'comb'
[pɛʁ]



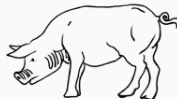
CVC
pêche 'peach'
[pɛʃ]



CV̆
pont 'bridge'
[pɔ̃]



CVN
pomme 'apple'
[pɔm]



CVC
porc 'pork'
[pɔʁ]



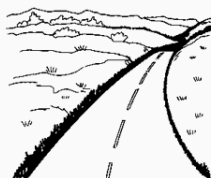
Filler
pelle 'shovel'
[pɛl]



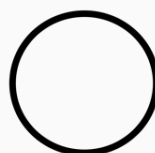
CVN
rhum 'rum'
[ʁəm]



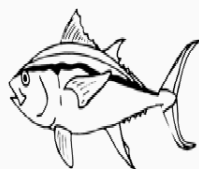
CVC
robe 'dress'
[ʁob]



Filler
route 'road'
[ʁut]



CṼ
rond 'circle'
[ʁɔ̃]



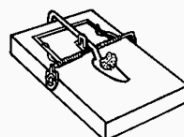
CṼ
thon 'tuna'
[tɔ̃]



CVN
tonne 'ton'
[tɔn]



CVC
torse 'torso'
[tɔʁs]



Filler
trappe 'trap'
[tʁap]



CVN
veine 'vein'
[vɛn]



CVC
verre 'glass'
[vɛʁ]



Filler
feu 'fire'
[fø]



CṼ
vin 'wine'
[vɛ̃]

ADDITIONAL SIGNIFICANT EFFECTS IN GAMMS

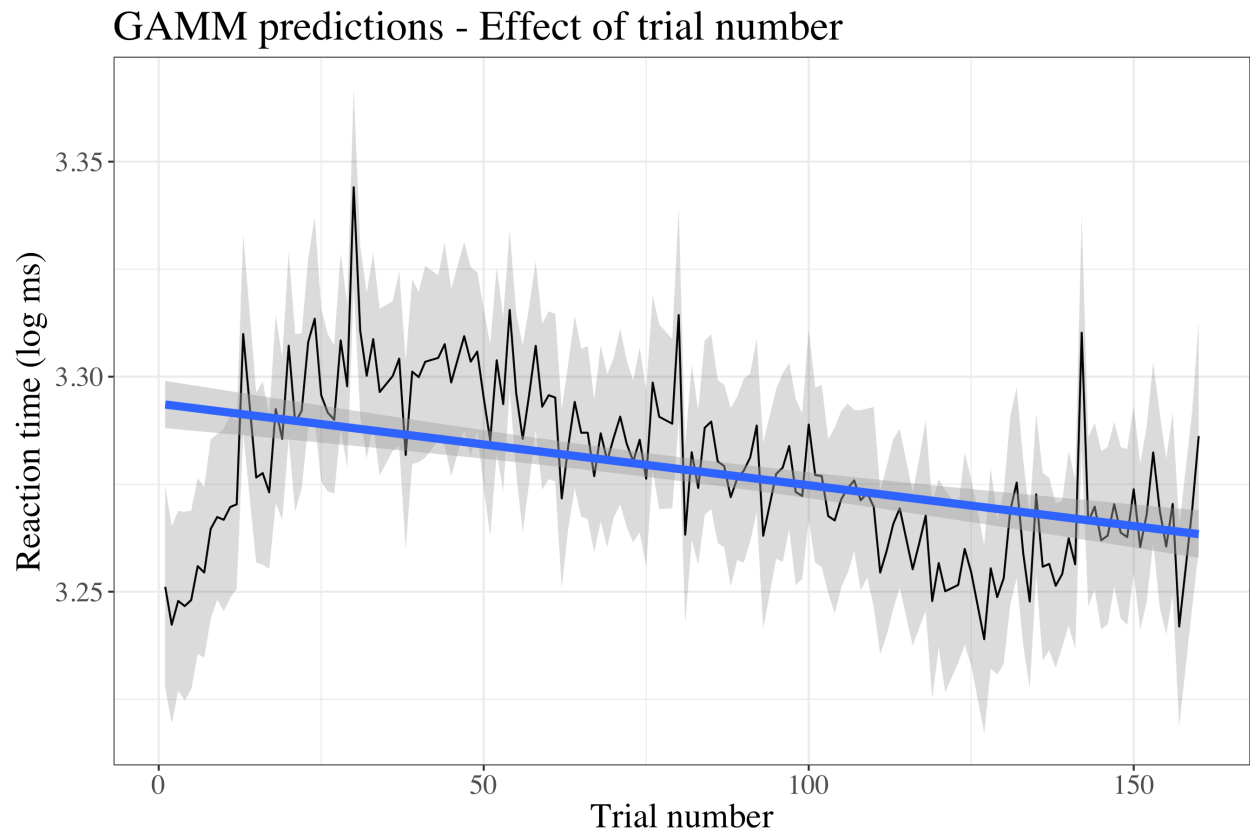


Figure A.1. Predicted data by TRIAL NUMBER – Exp. 1a RTs

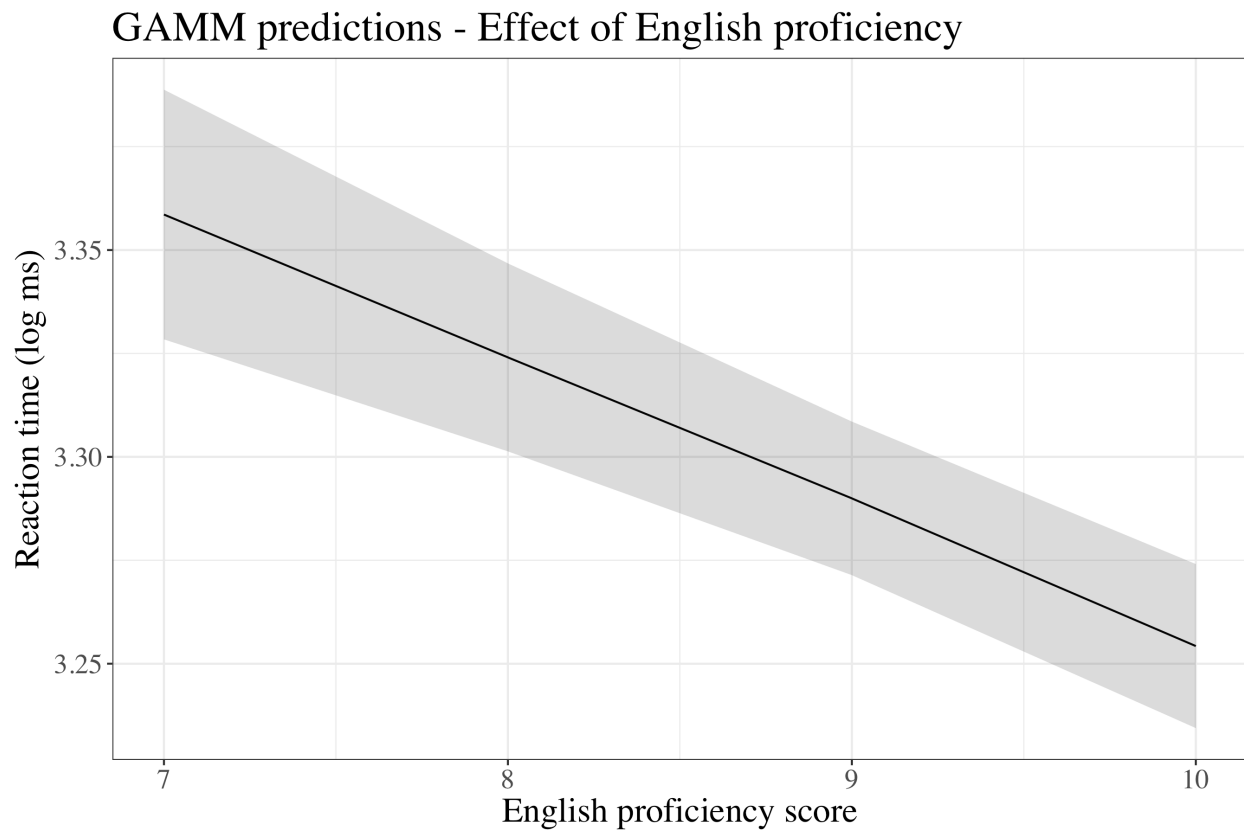


Figure A.2. Predicted data by L2 (ENGLISH) PROFICIENCY score – Exp. 1a RTs

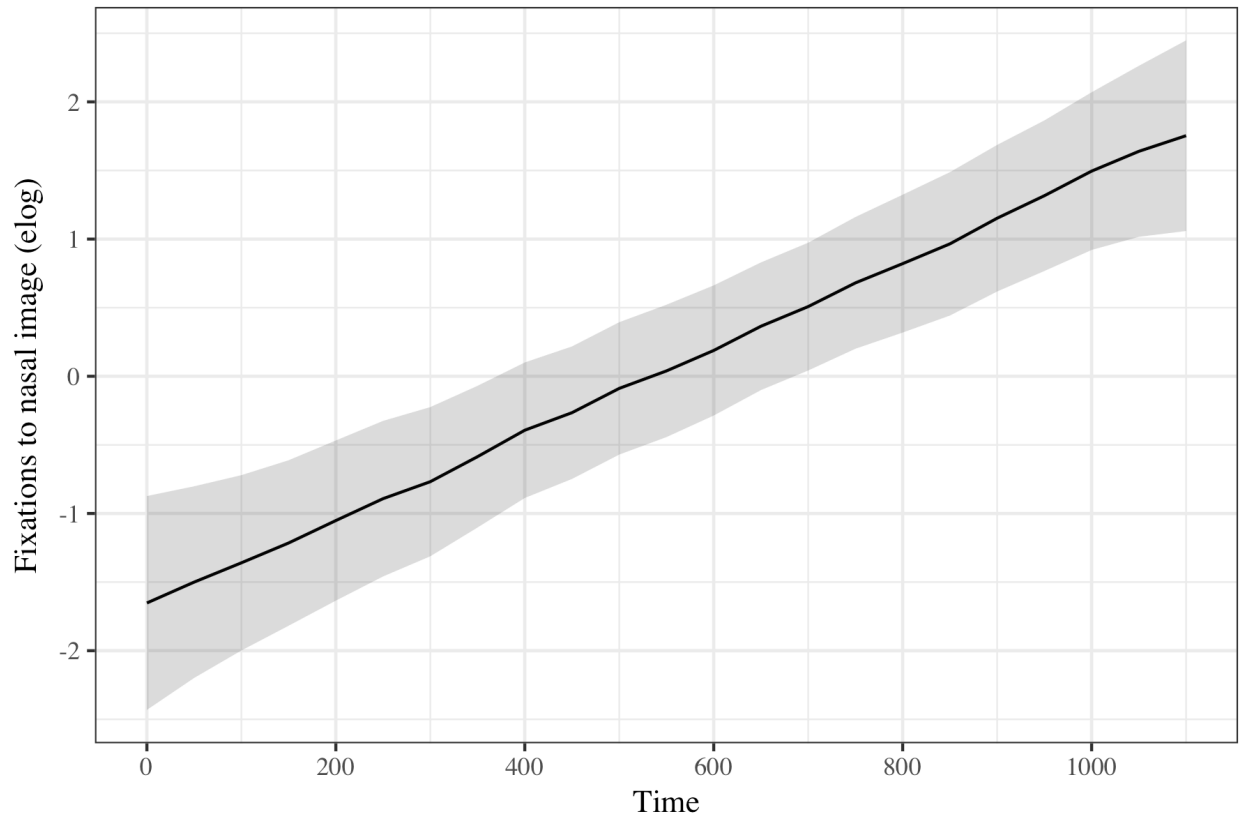


Figure A.3. Predicted data through TIME – Exp. 1a elog %fix to $C\tilde{V}$, $C\tilde{V}$ choices

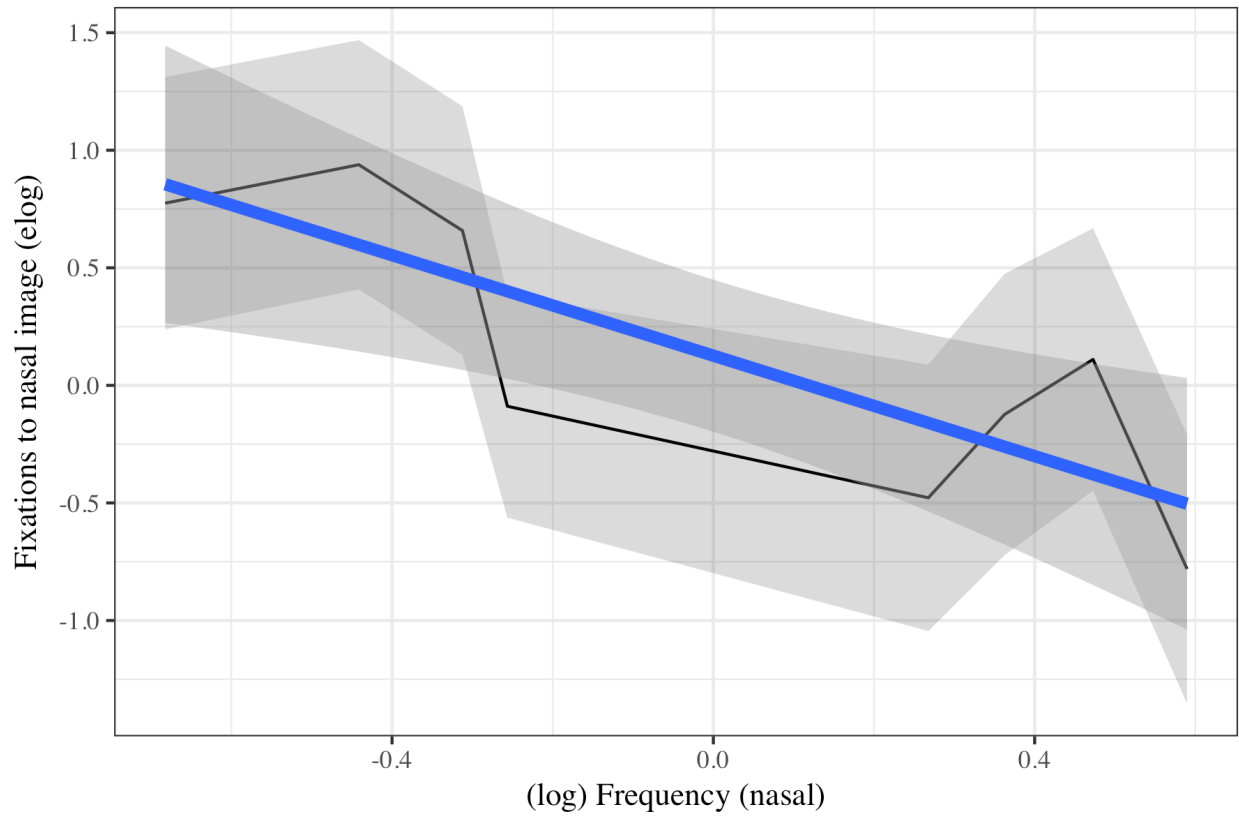


Figure A.4. Predicted data by FREQUENCY (NASAL) – Exp. 1a elog %fix to $C\tilde{V}$, $C\tilde{V}$ choices

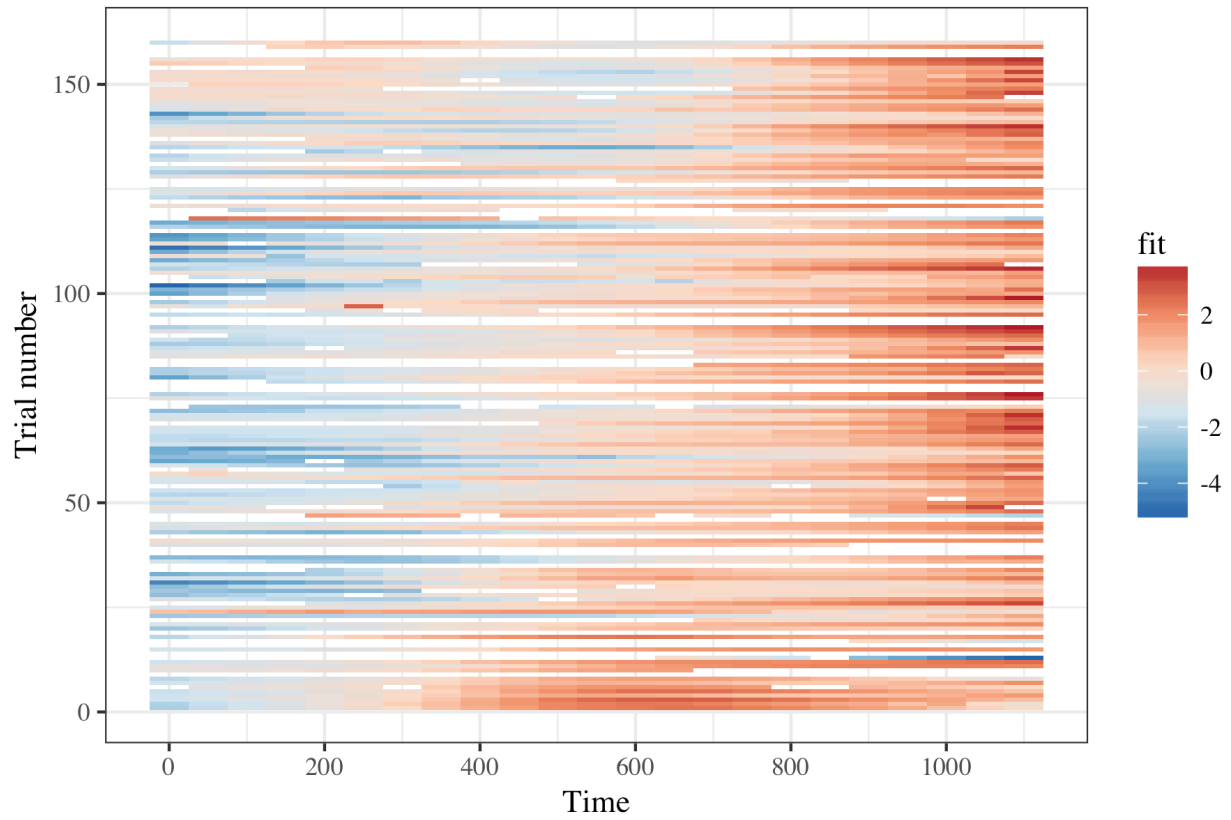


Figure A.5. Predicted data by TRIAL NUMBER through TIME – Exp. 1a elog %fix to $C\tilde{V}$, $C\tilde{V}$ choices

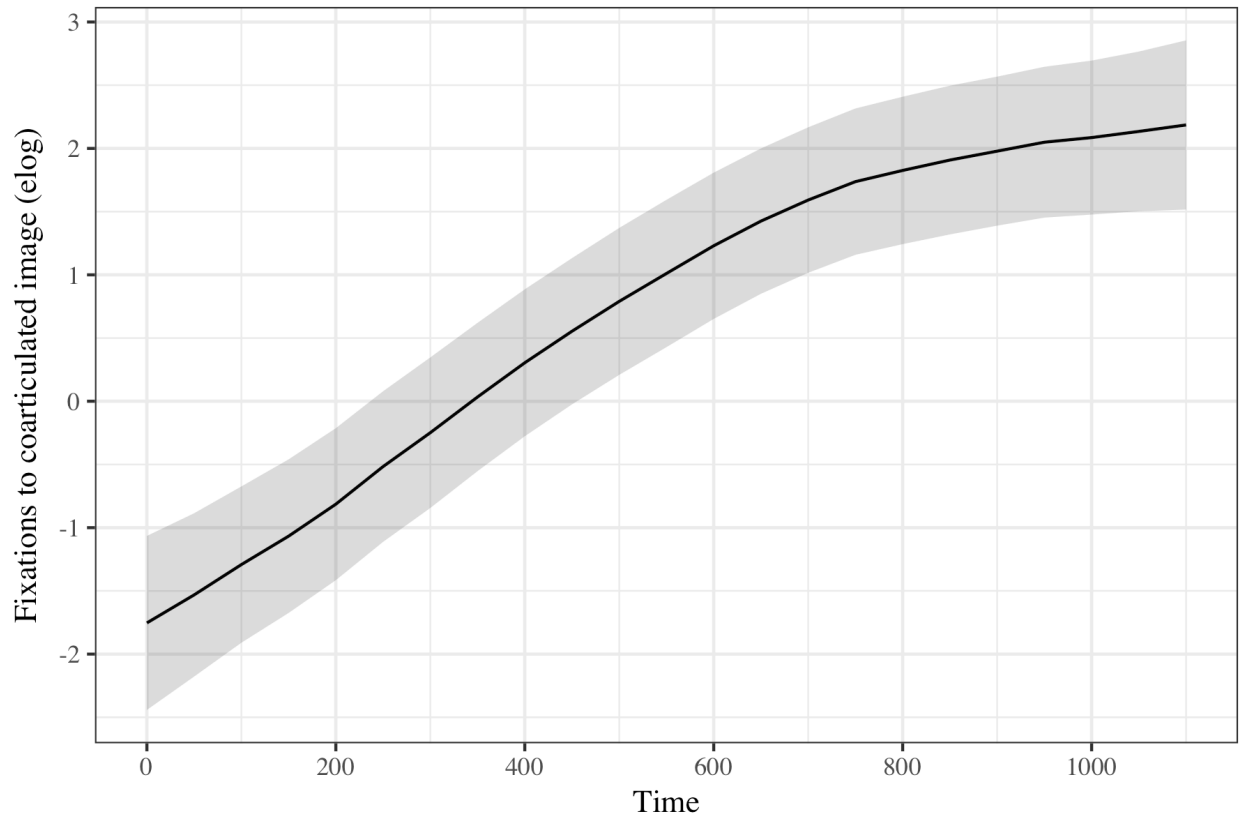


Figure A.6. Predicted data through TIME – Exp. 1a elog %fix to CVN, CVN choices

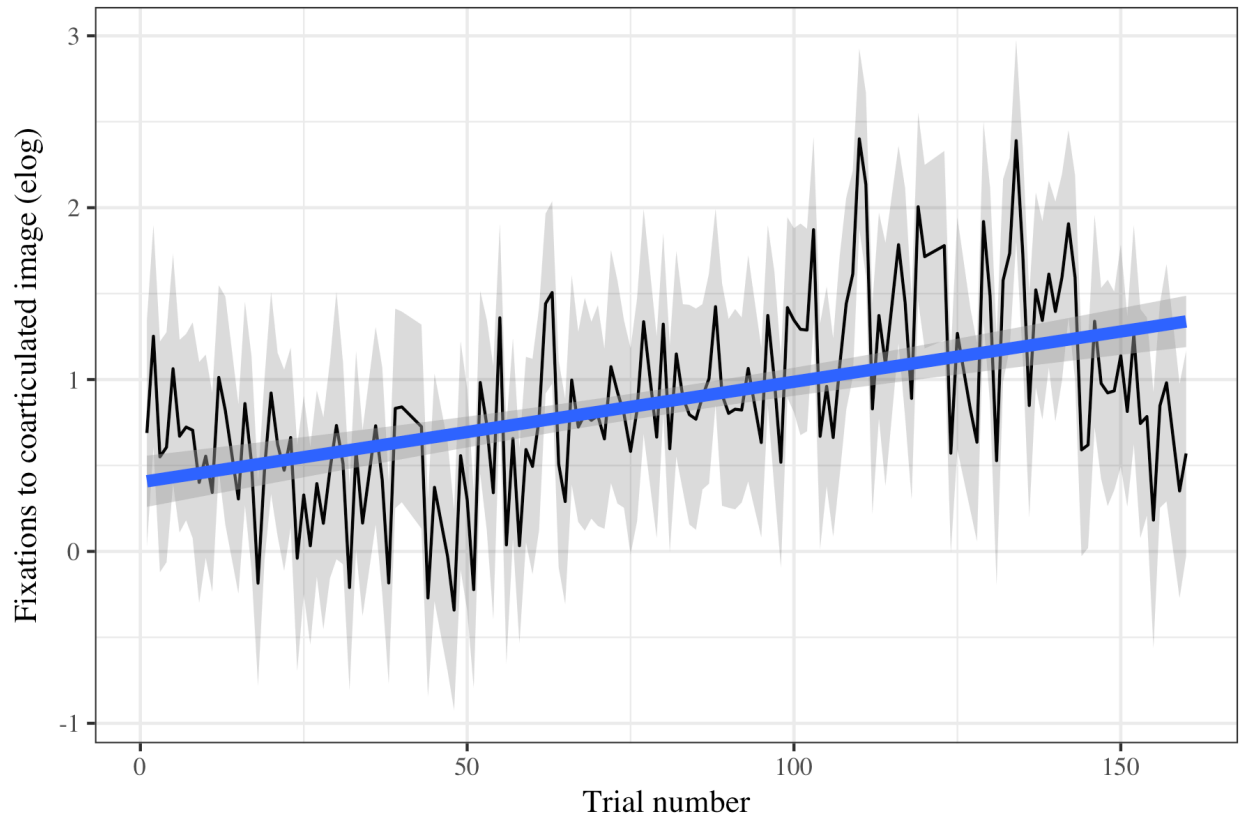


Figure A.7. Predicted data by TRIAL NUMBER – Exp. 1a elog %fix to CVN, CVN choices

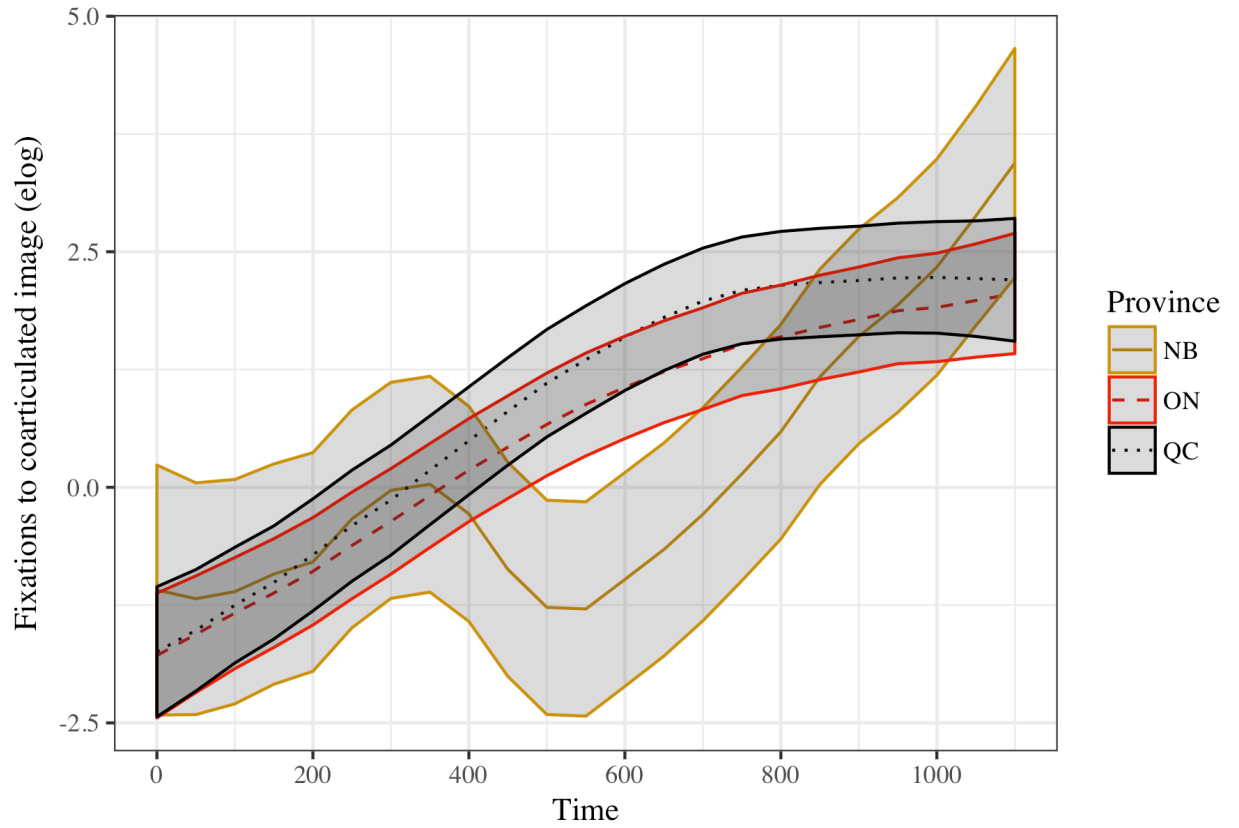


Figure A.8. Predicted data by PROVINCE OF ORIGIN through TIME – Exp. 1a elog %fix to CVN, CVN choices

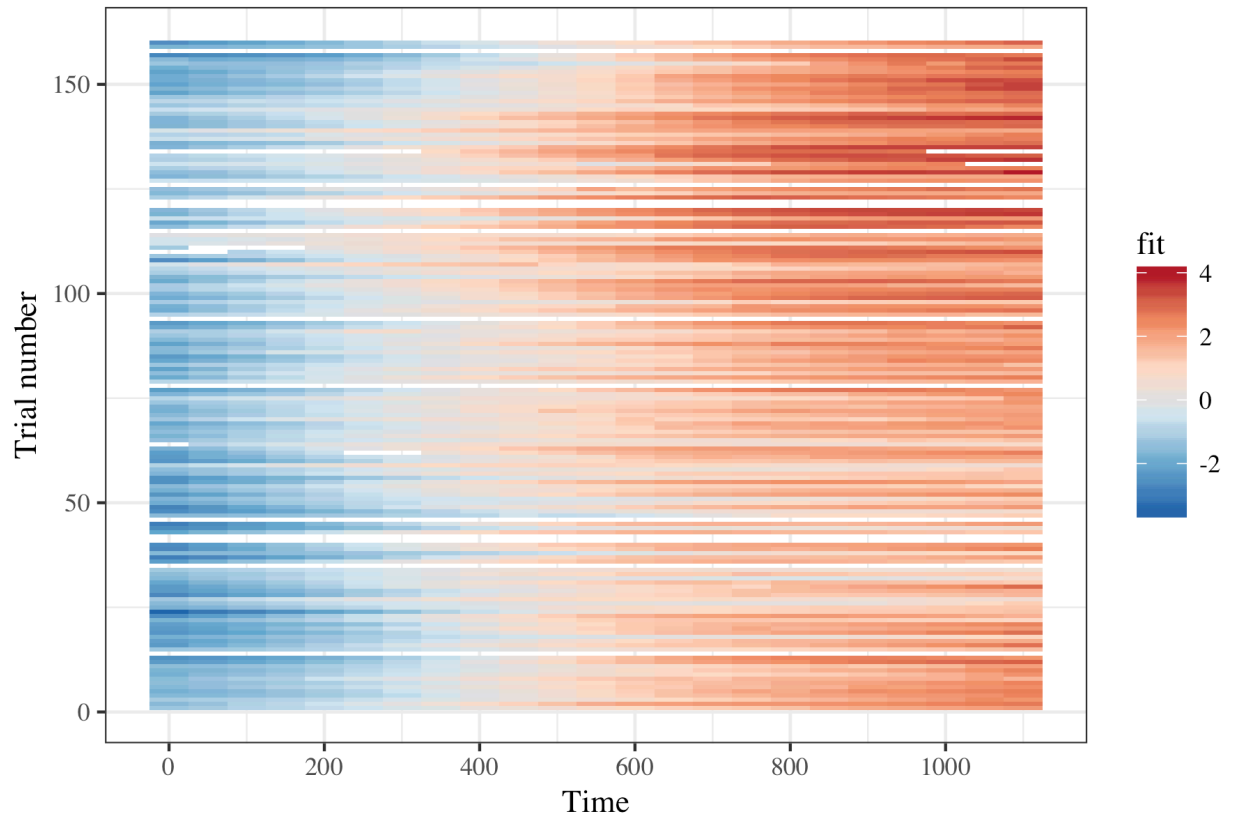


Figure A.9. Predicted data by TRIAL NUMBER through TIME – Exp. 1a elog %fix to CVN, CVN choices

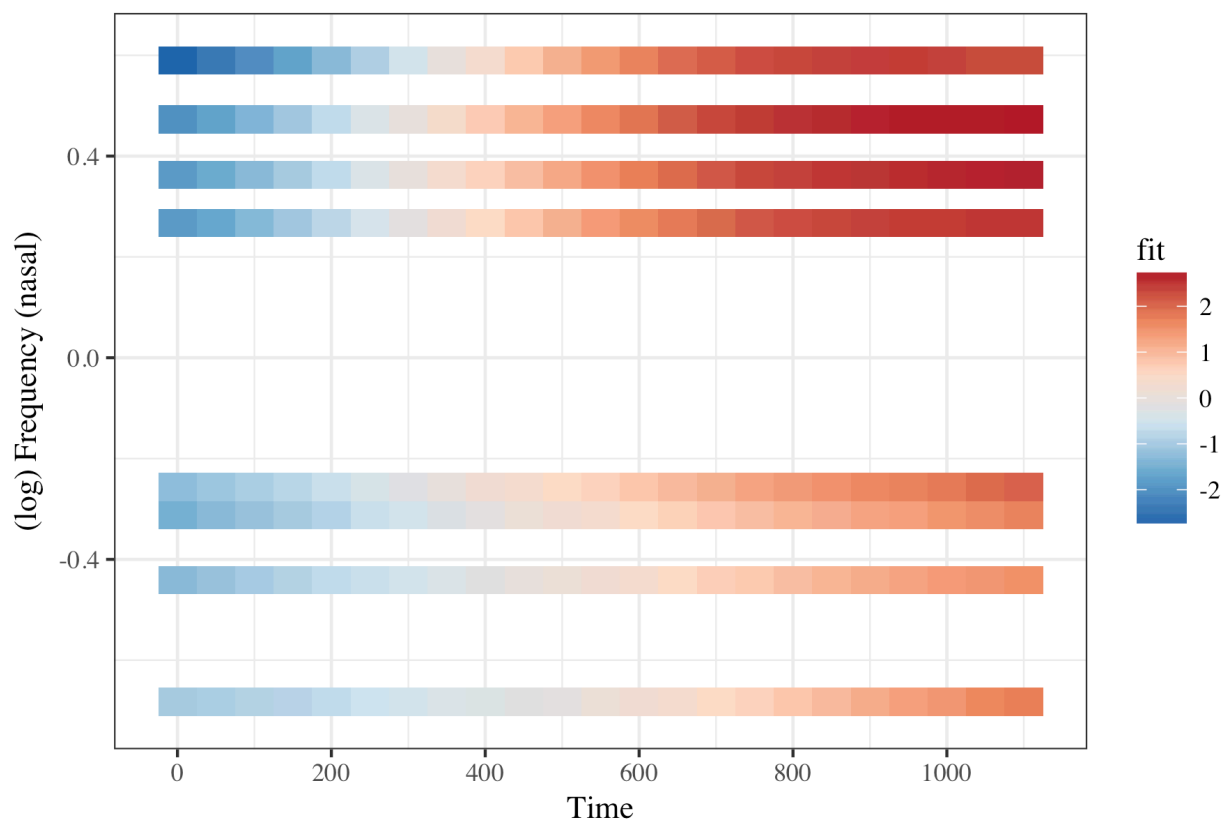


Figure A.10. Predicted data by FREQUENCY (NASAL) through TIME – Exp. 1a elog %fix to CVN, CVN choices

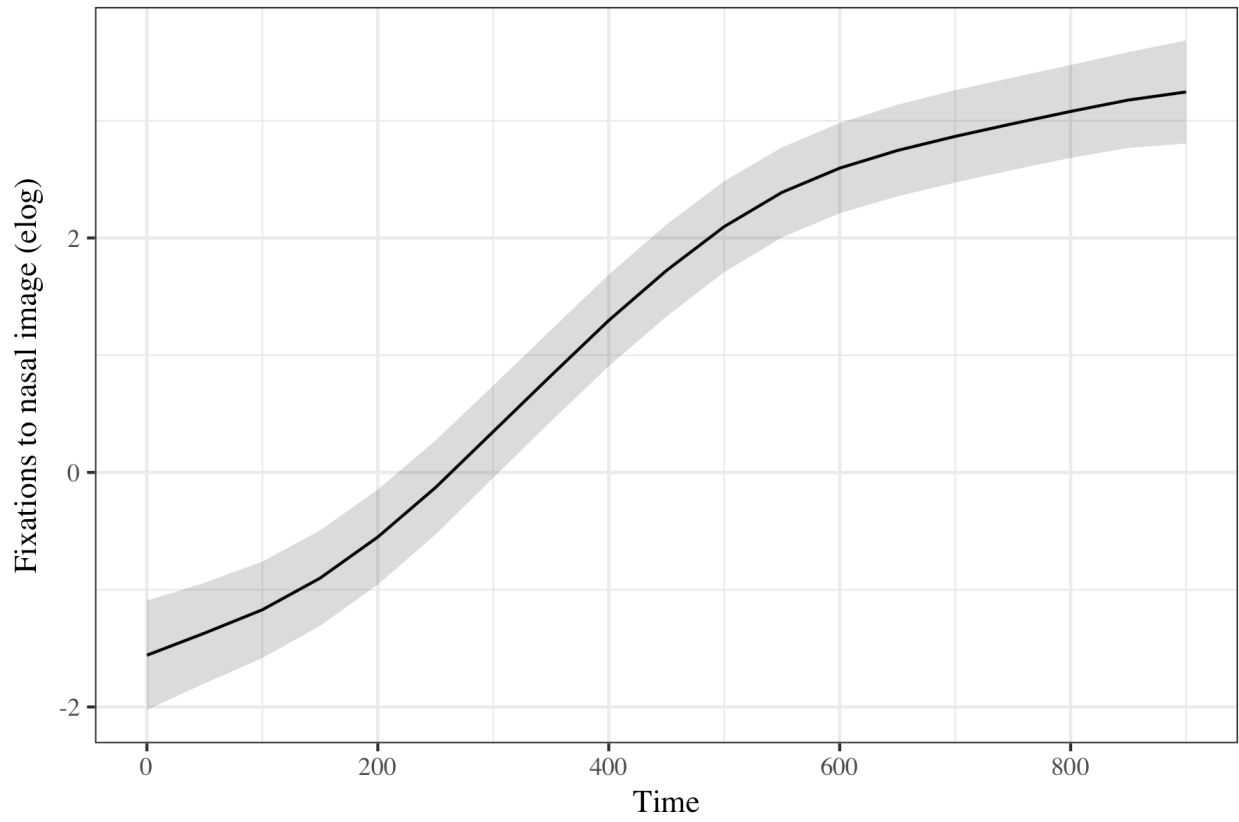


Figure A.II. Predicted data through TIME – Exp. 1b elog %fix to $C\tilde{V}$, $C\tilde{V}$ choices

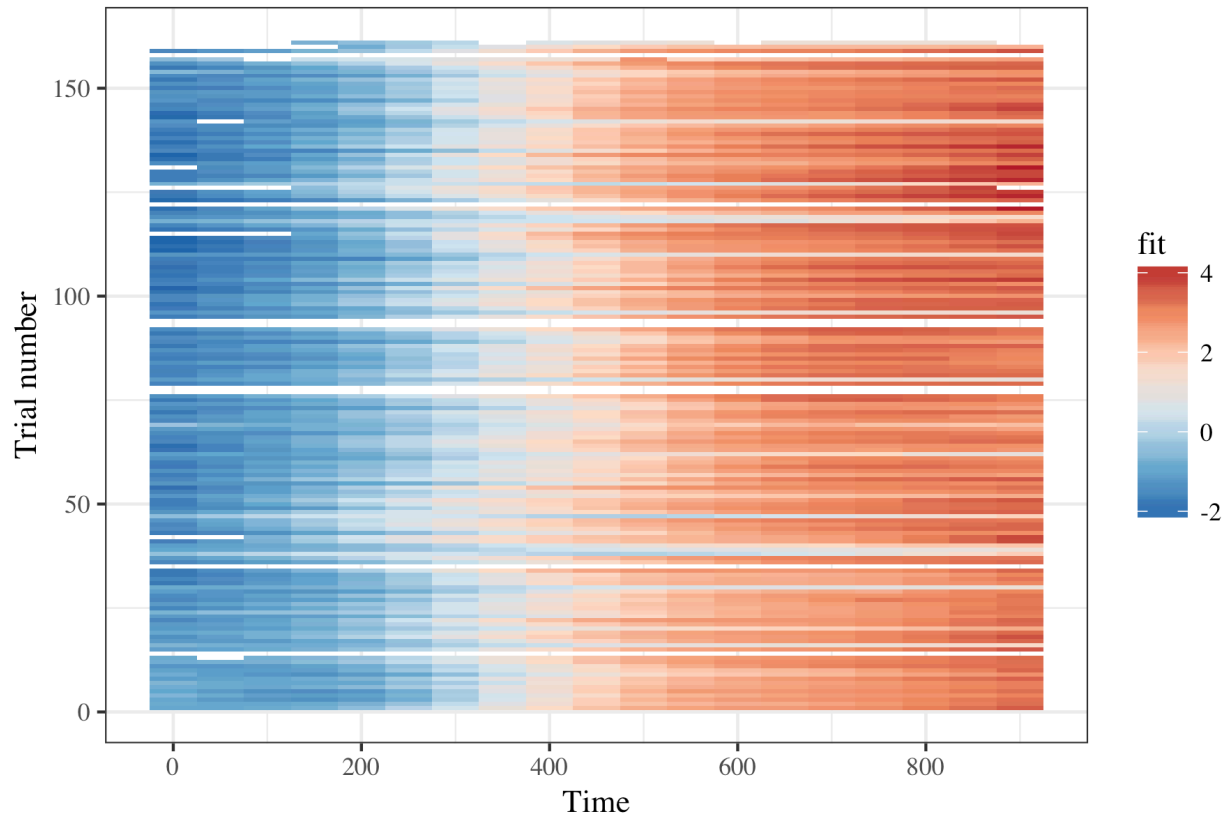


Figure A.12. Predicted data by TRIAL NUMBER through TIME – Exp. 1b elog %fix to $C\tilde{V}$, $C\tilde{V}$ choices

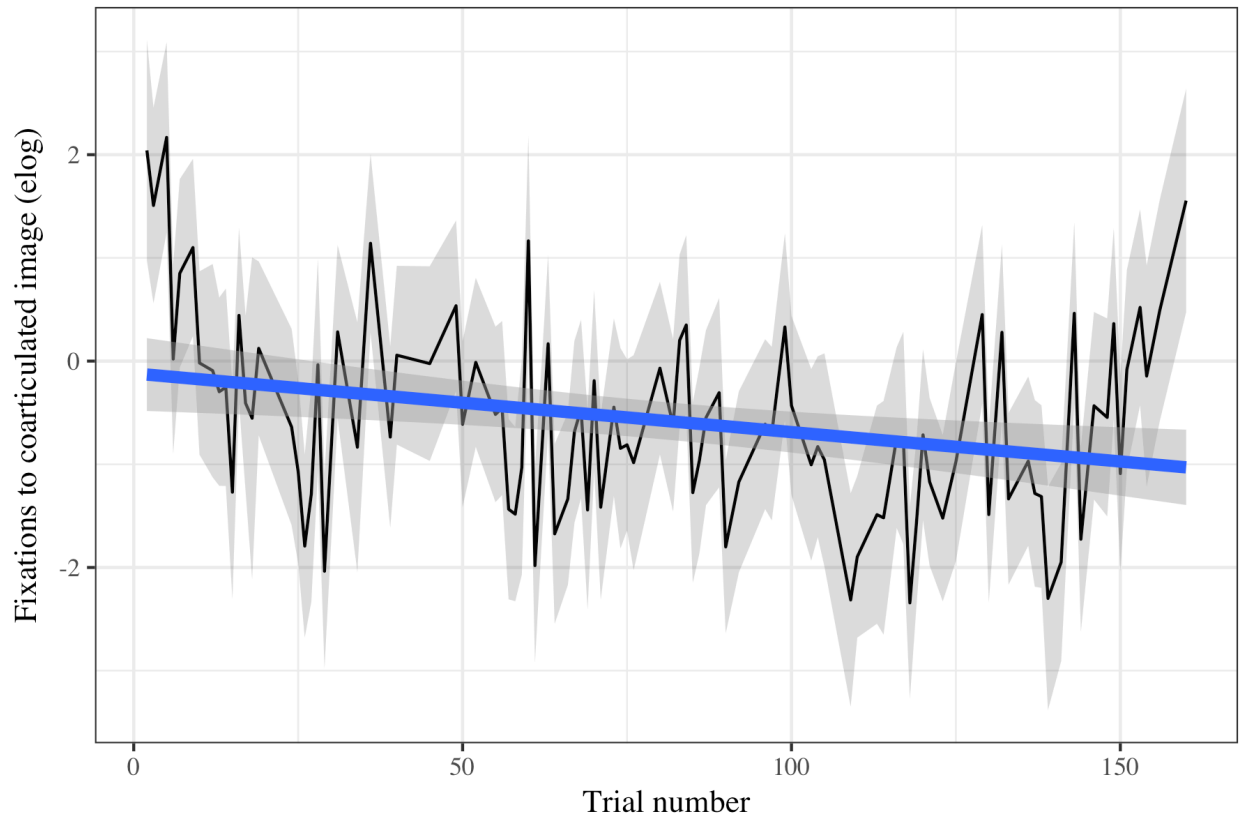


Figure A.13. Predicted data by TRIAL NUMBER – Exp. 1b elog %fix to CVN, CVN choices

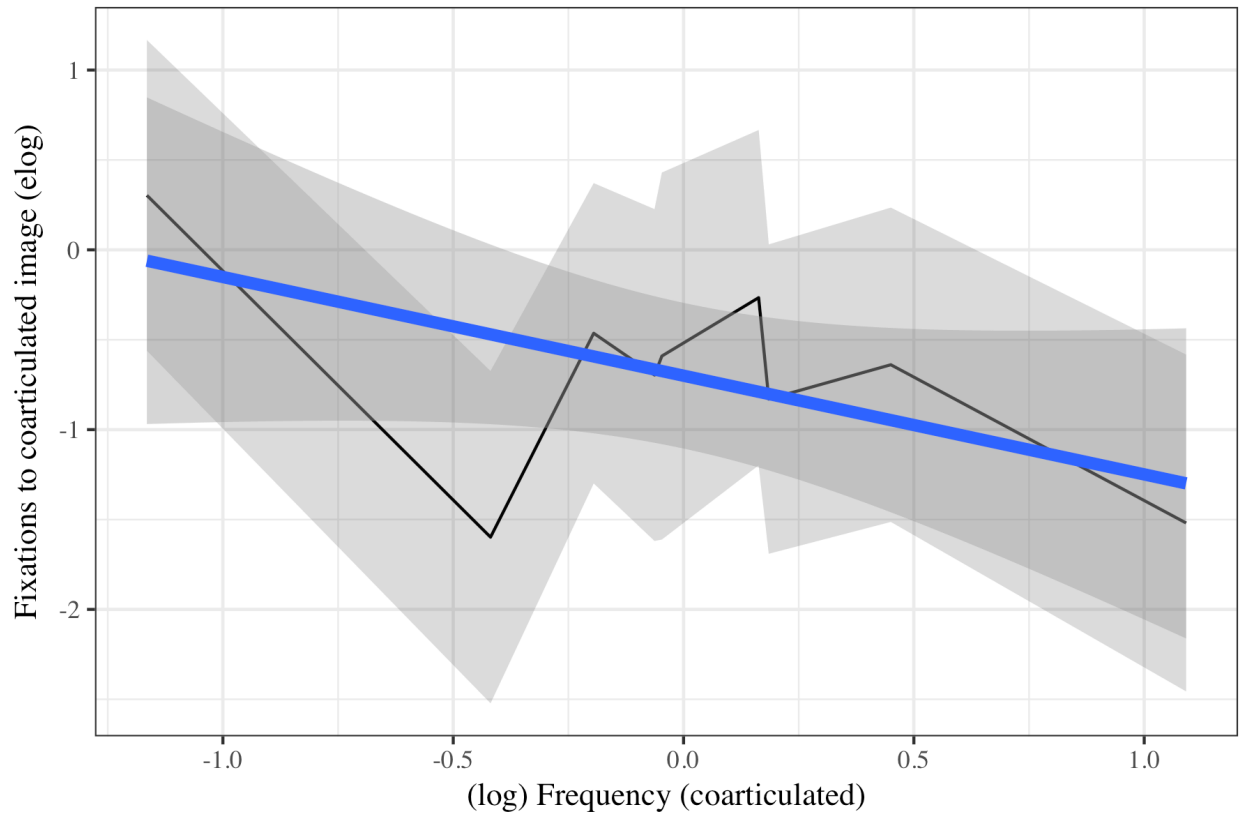


Figure A.14. Predicted data by FREQUENCY (COARTICULATED) – Exp. 1b elog %fix to CVN, CVN choices

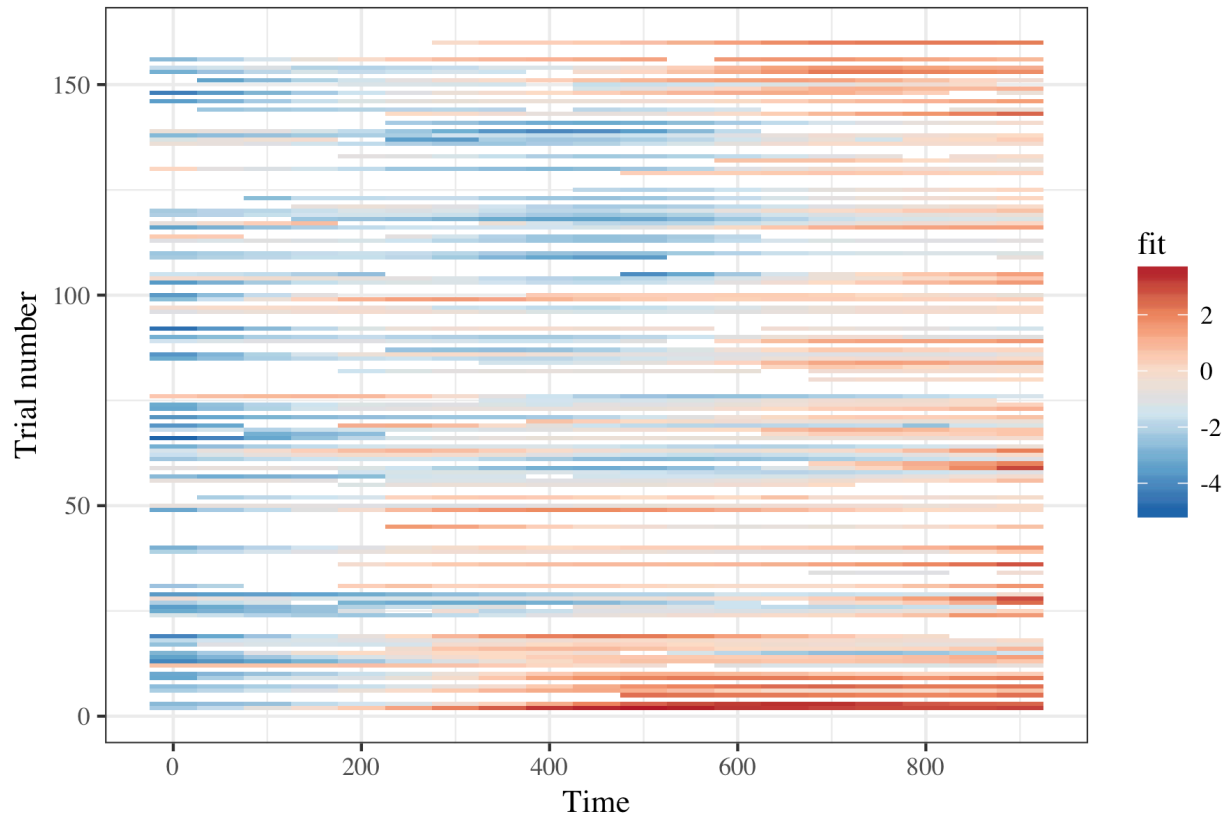


Figure A.15. Predicted data by TRIAL NUMBER – Exp. 1b elog %fix to CVN, CVN choices

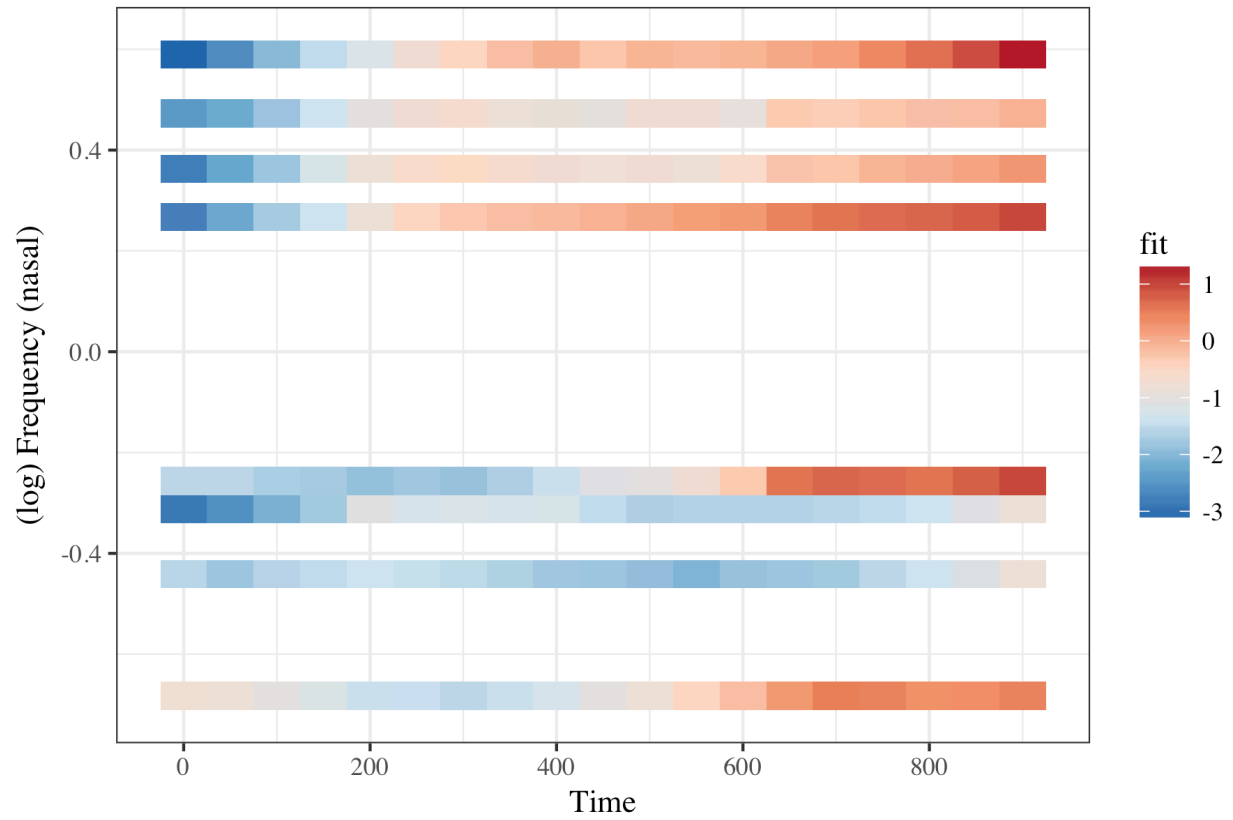


Figure A.16. Predicted data by TRIAL NUMBER – Exp. 1b elog %fix to CVN, CVN choices

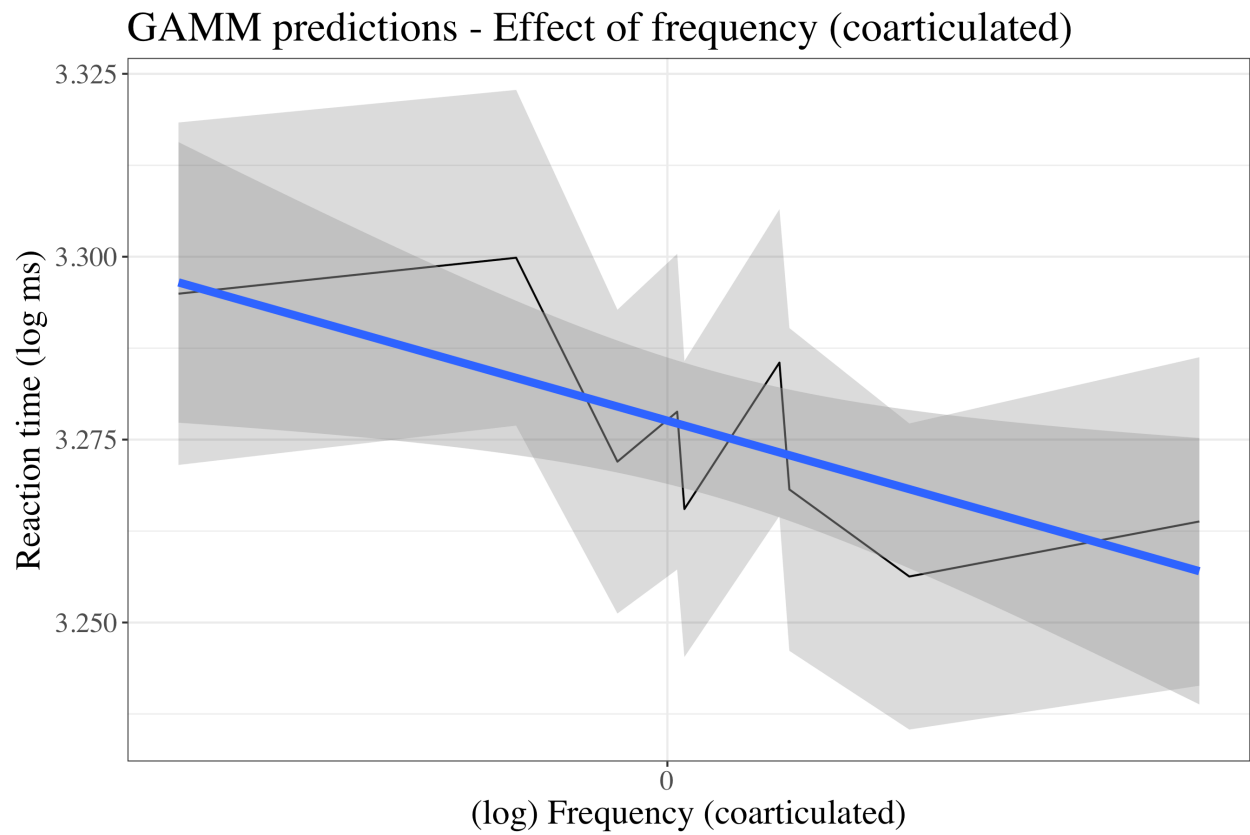


Figure A.17. Predicted data by (LOG) FREQUENCY (COARTICULATED) – Exp. 2a RTs

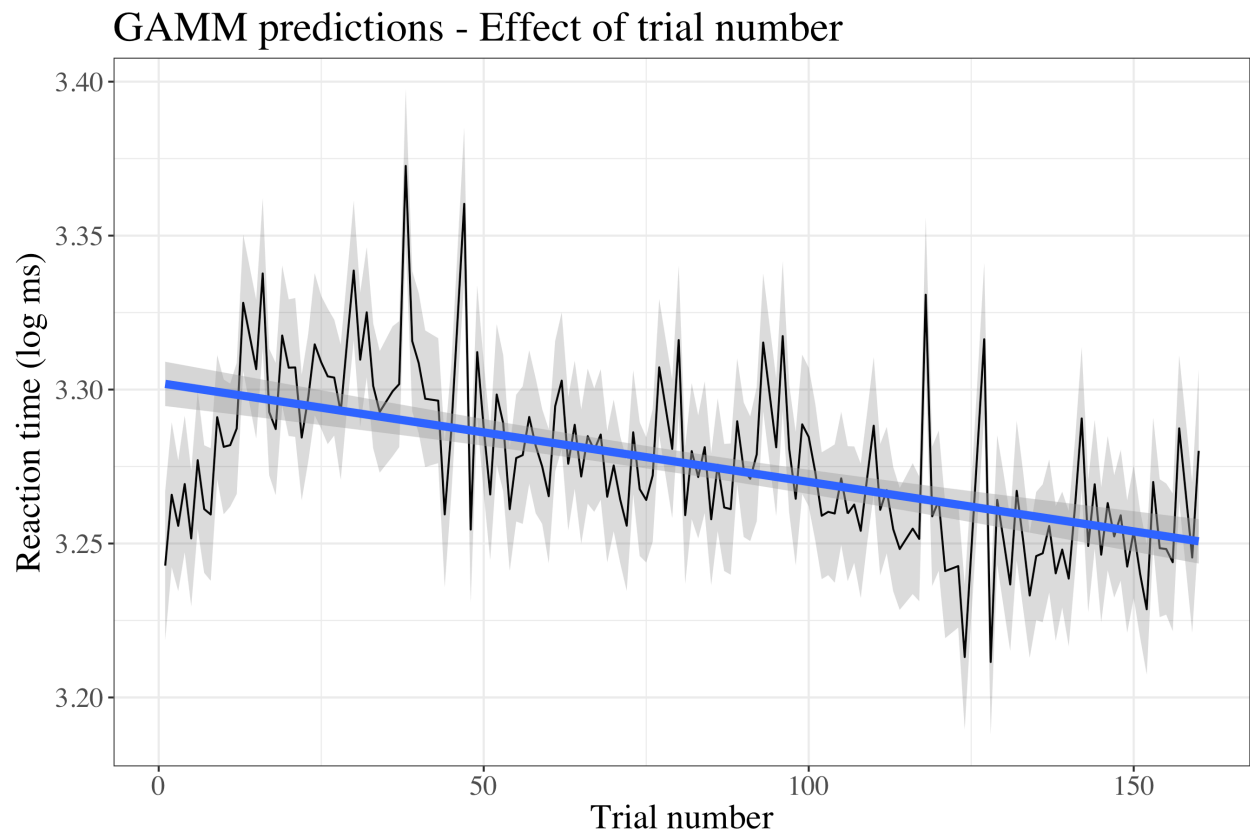


Figure A.18. Predicted data by TRIAL NUMBER – Exp. 2a RTs

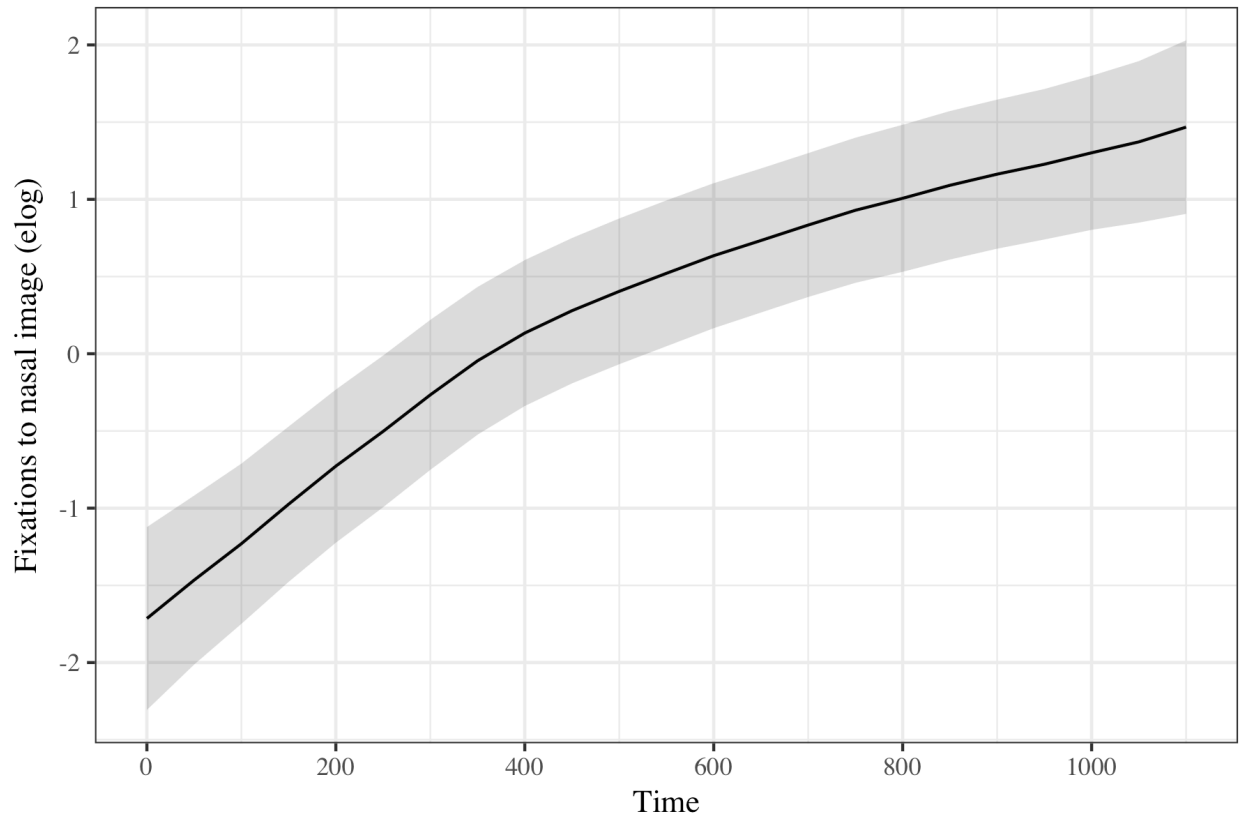


Figure A.19. Predicted data through TIME – Exp. 2a elog %fix to $C\tilde{V}$, $C\tilde{V}$ choices

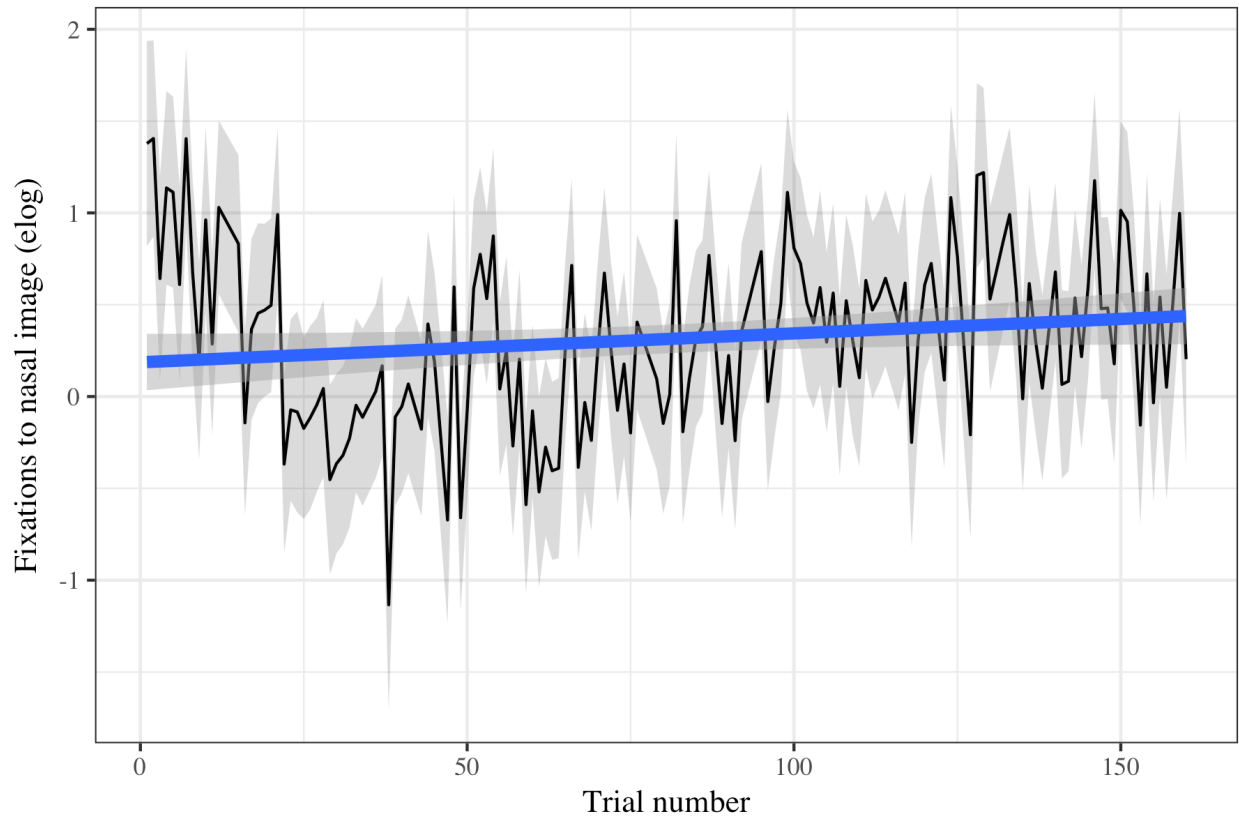


Figure A.20. Predicted data by TRIAL NUMBER – Exp. 2a elog %fix to $C\tilde{V}$, $C\tilde{V}$ choices

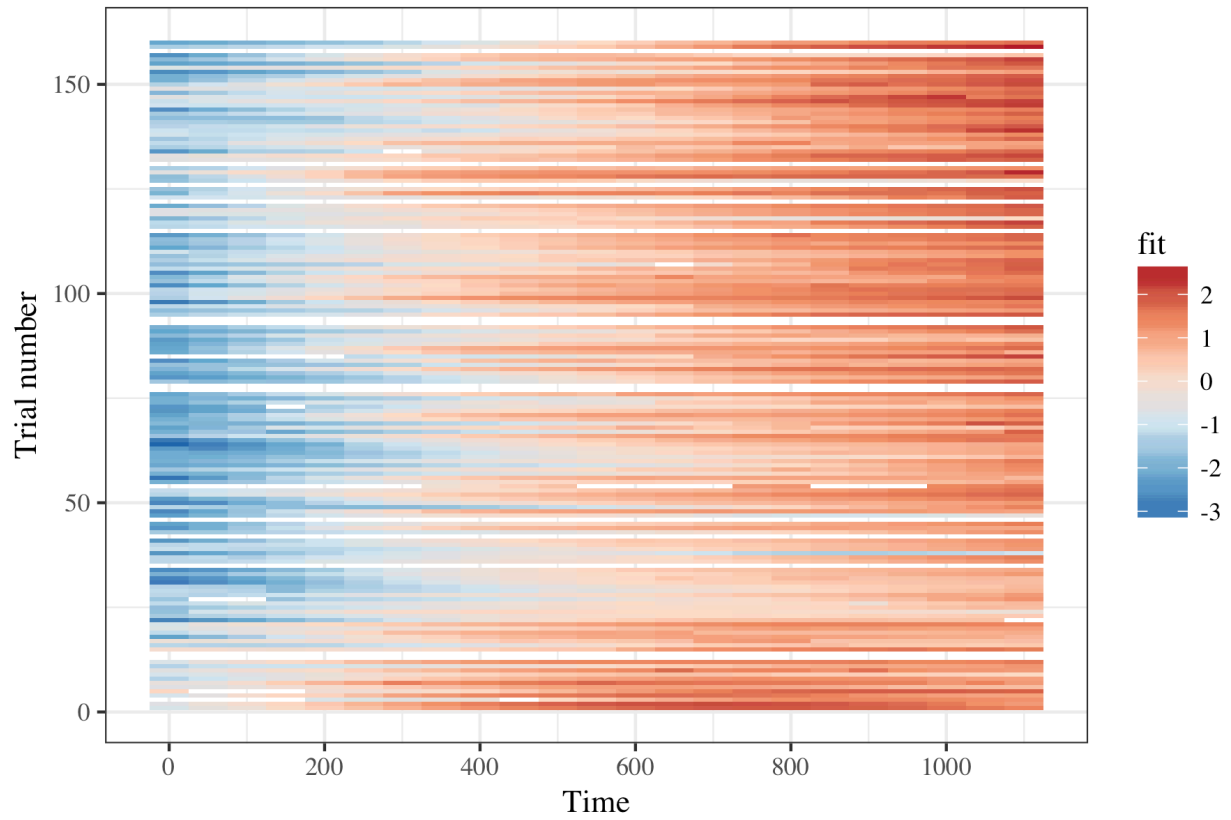


Figure A.21. Predicted data by TRIAL NUMBER through TIME – Exp. 2a elog %fix to $C\tilde{V}$, $C\tilde{V}$ choices

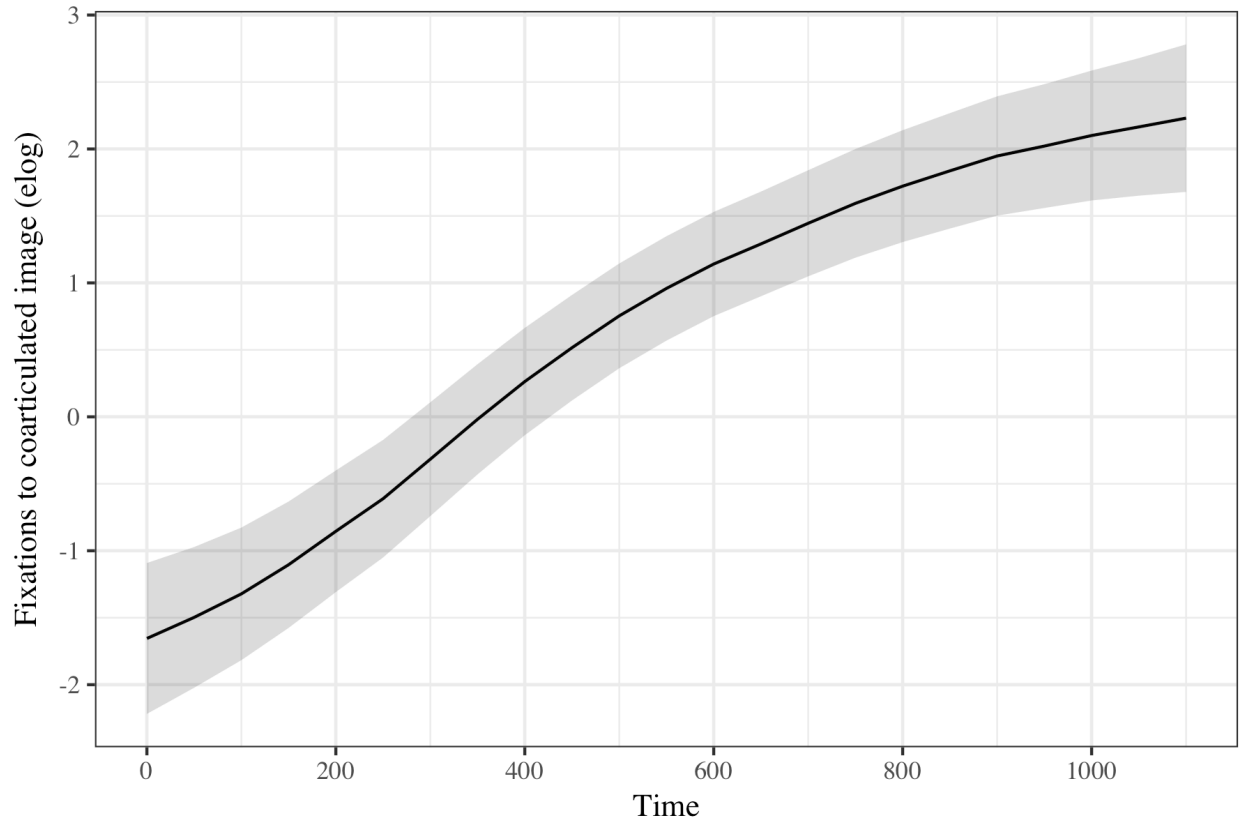


Figure A.22. Predicted data through TIME – Exp. 2a elog %fix to CVN, CVN choices

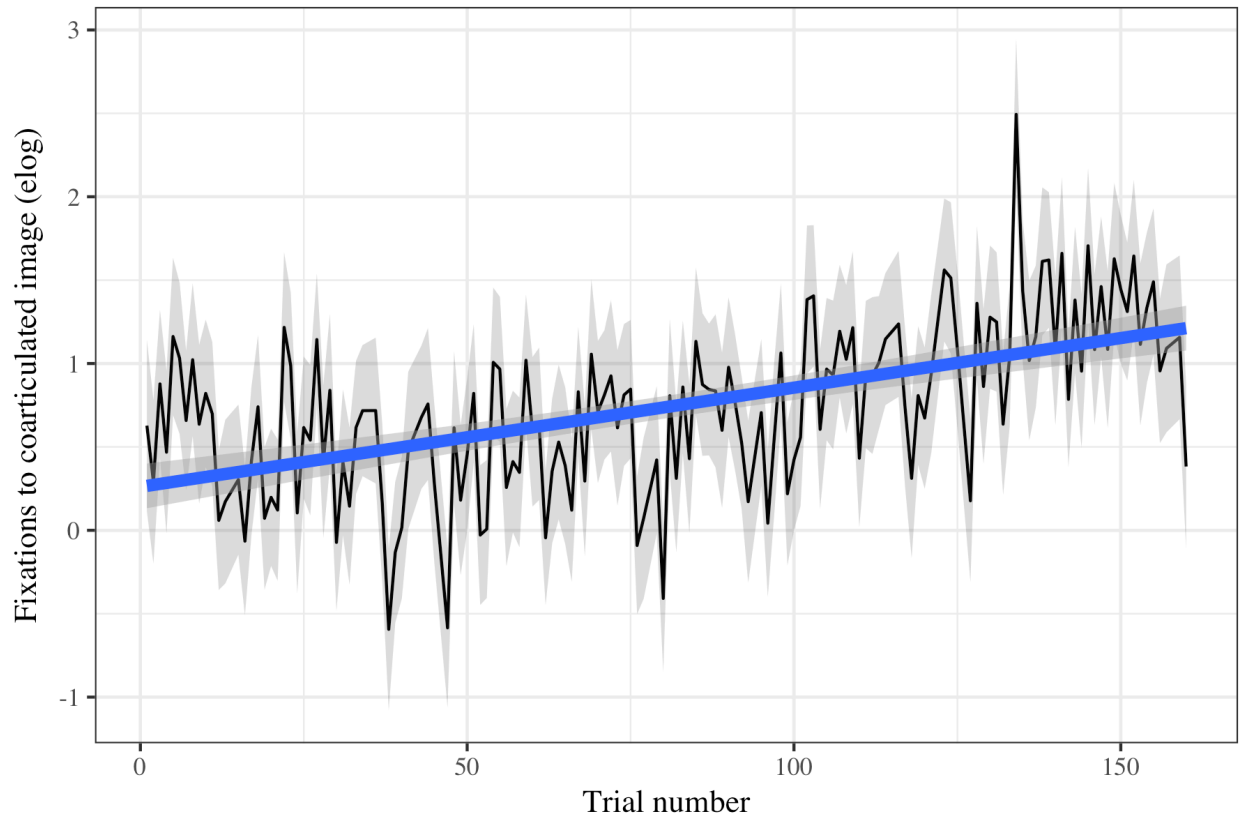


Figure A.23. Predicted data by TRIAL NUMBER – Exp. 2a elog %fix to CVN, CVN choices

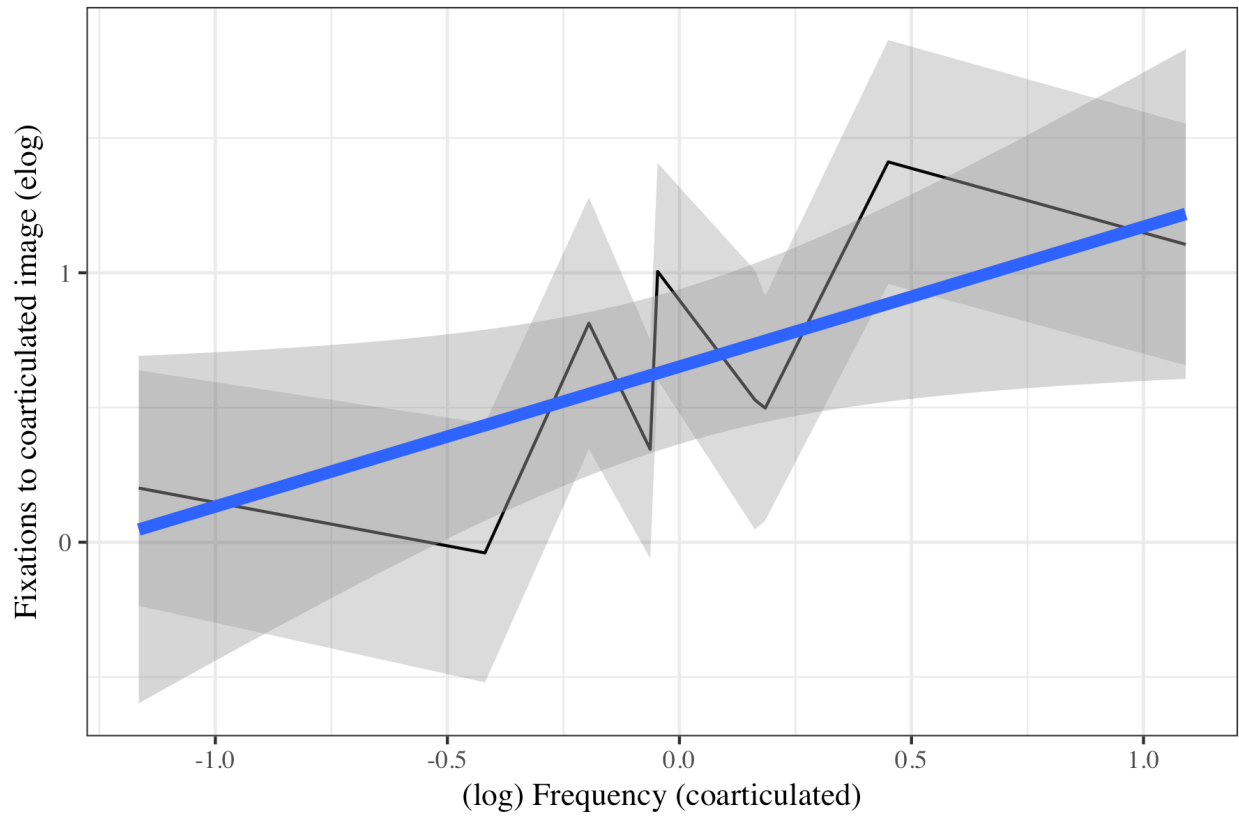


Figure A.24. Predicted data by FREQUENCY (COARTICULATED) – Exp. 2a elog %fix to CVN, CVN choices

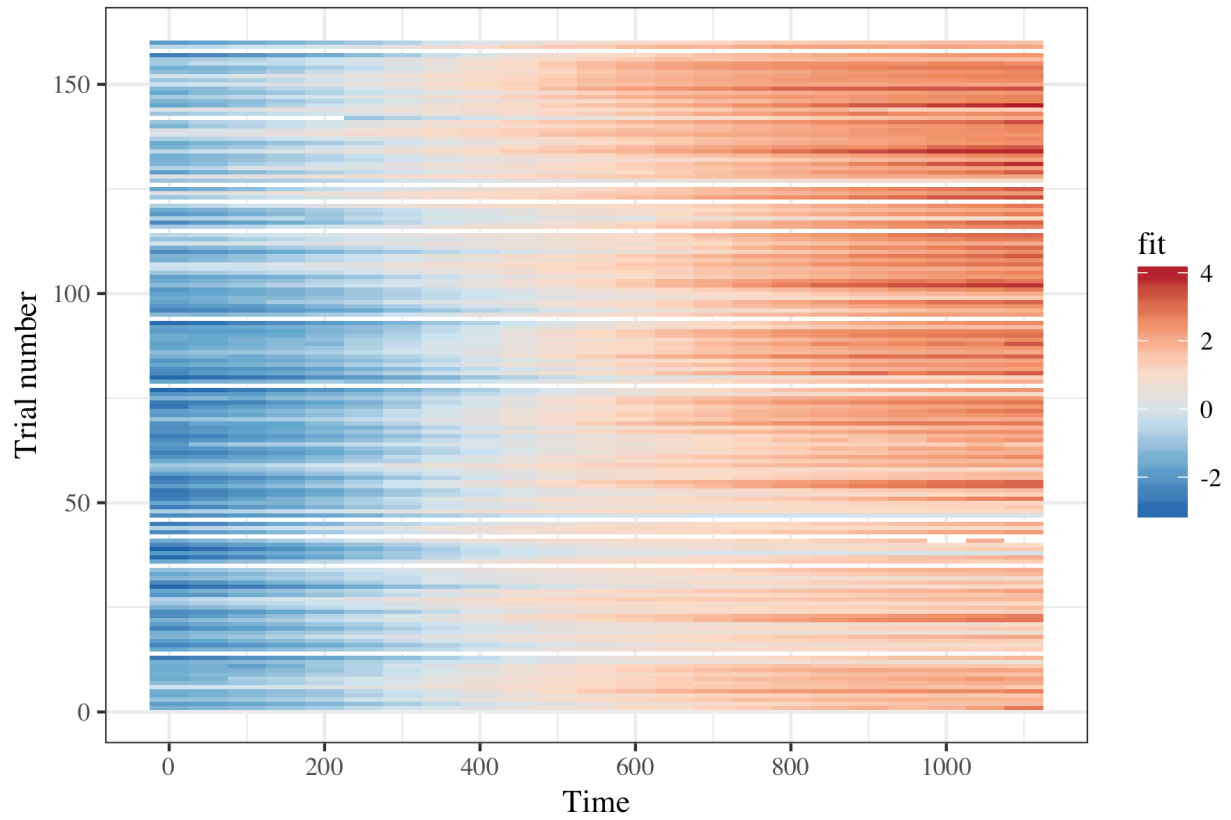


Figure A.25. Predicted data by TRIAL NUMBER through TIME – Exp. 2a elog %fix to CVN, CVN choices

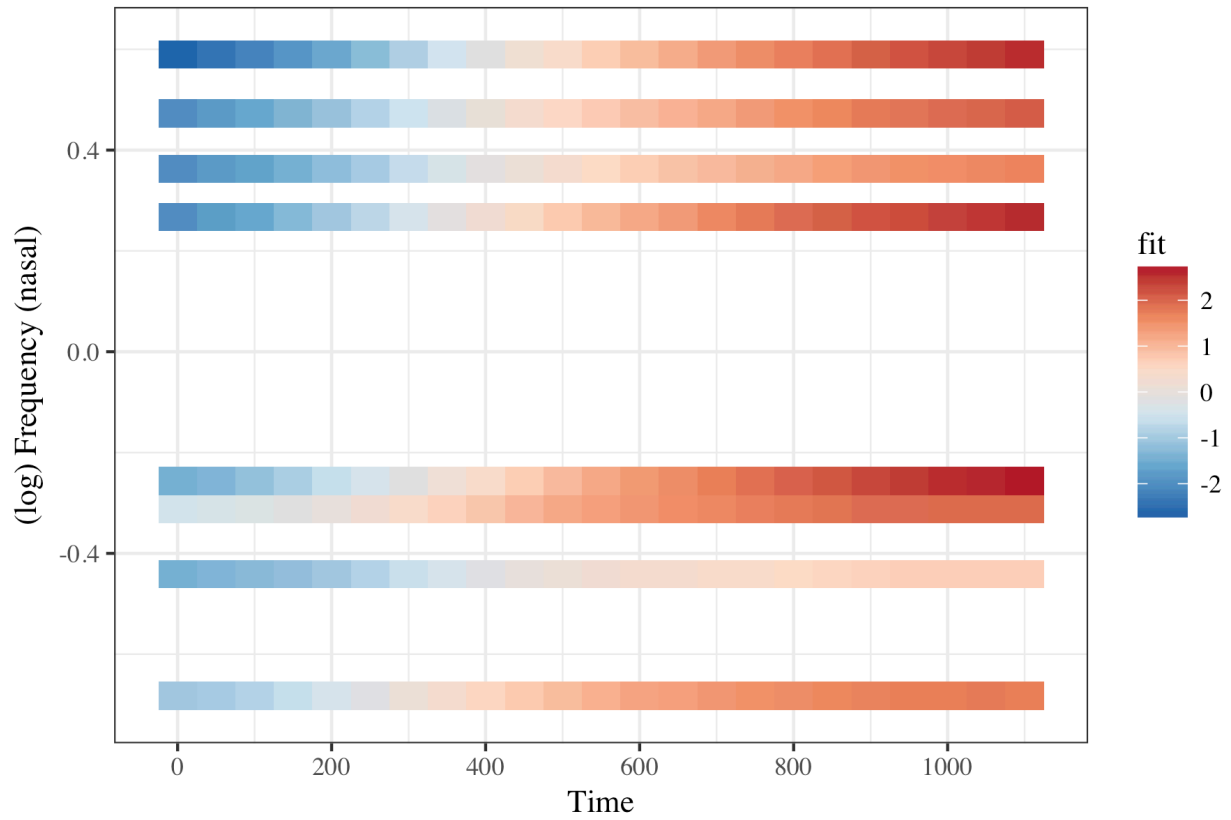


Figure A.26. Predicted data by FREQUENCY (NASAL) through TIME – Exp. 2a elog %fix to CVN, CVN choices

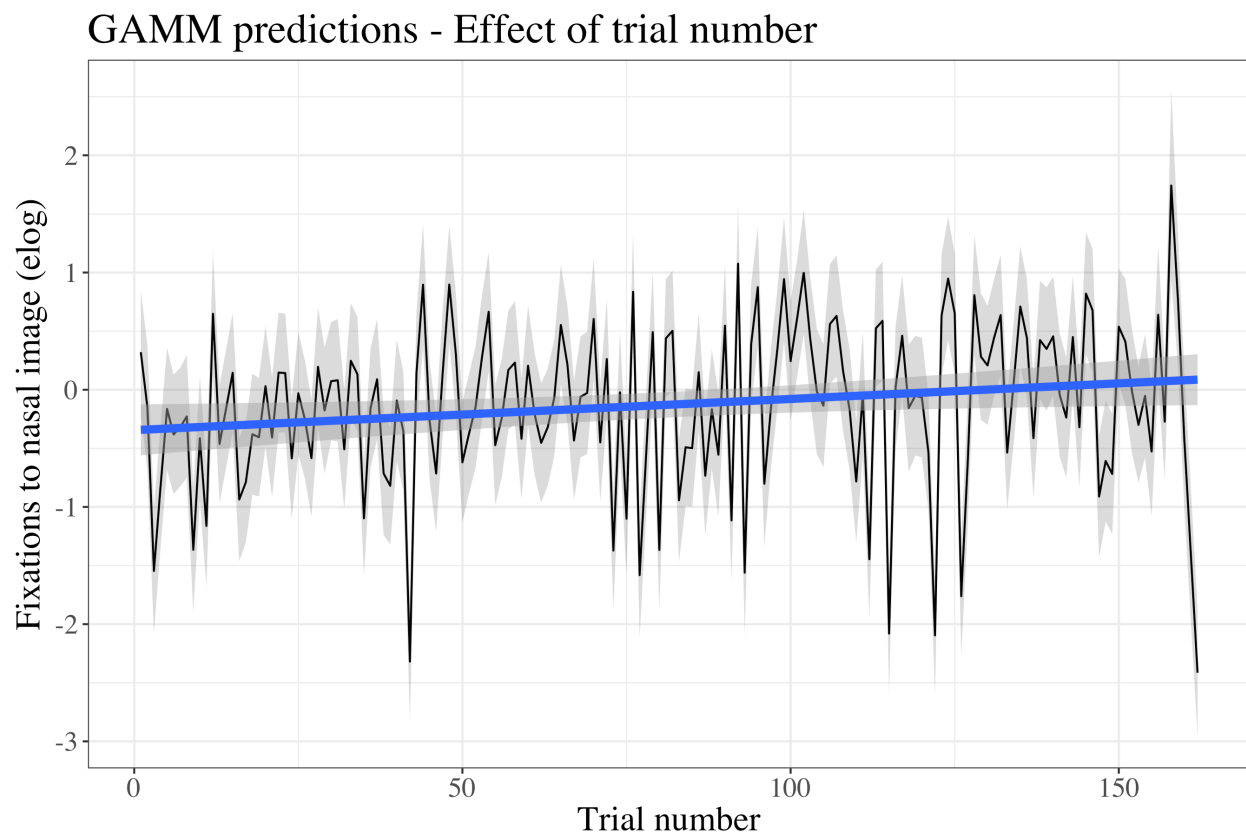


Figure A.27. Predicted data by TRIAL NUMBER – Exp. 2b RTs

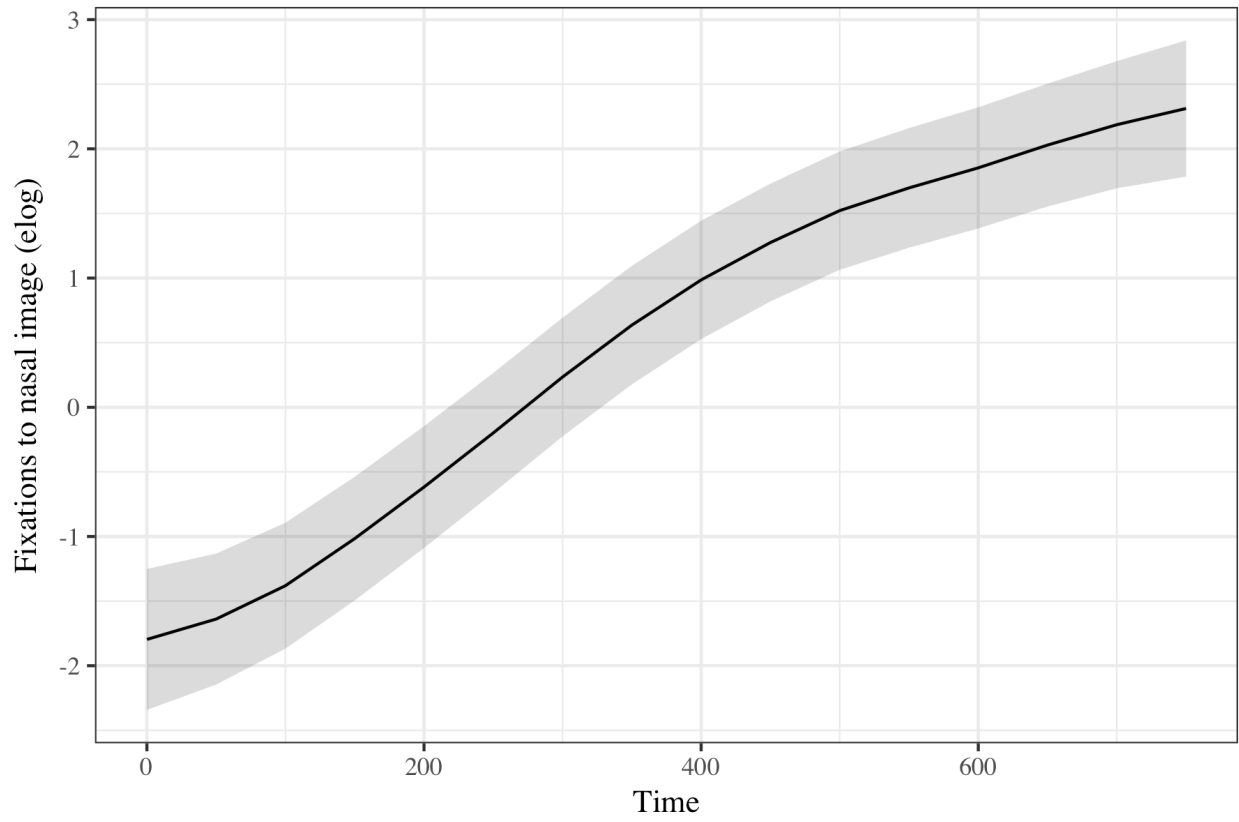


Figure A.28. Predicted data by TRIAL NUMBER – Exp. 2b elog %fix to $C\tilde{V}$, $C\tilde{V}$ choices

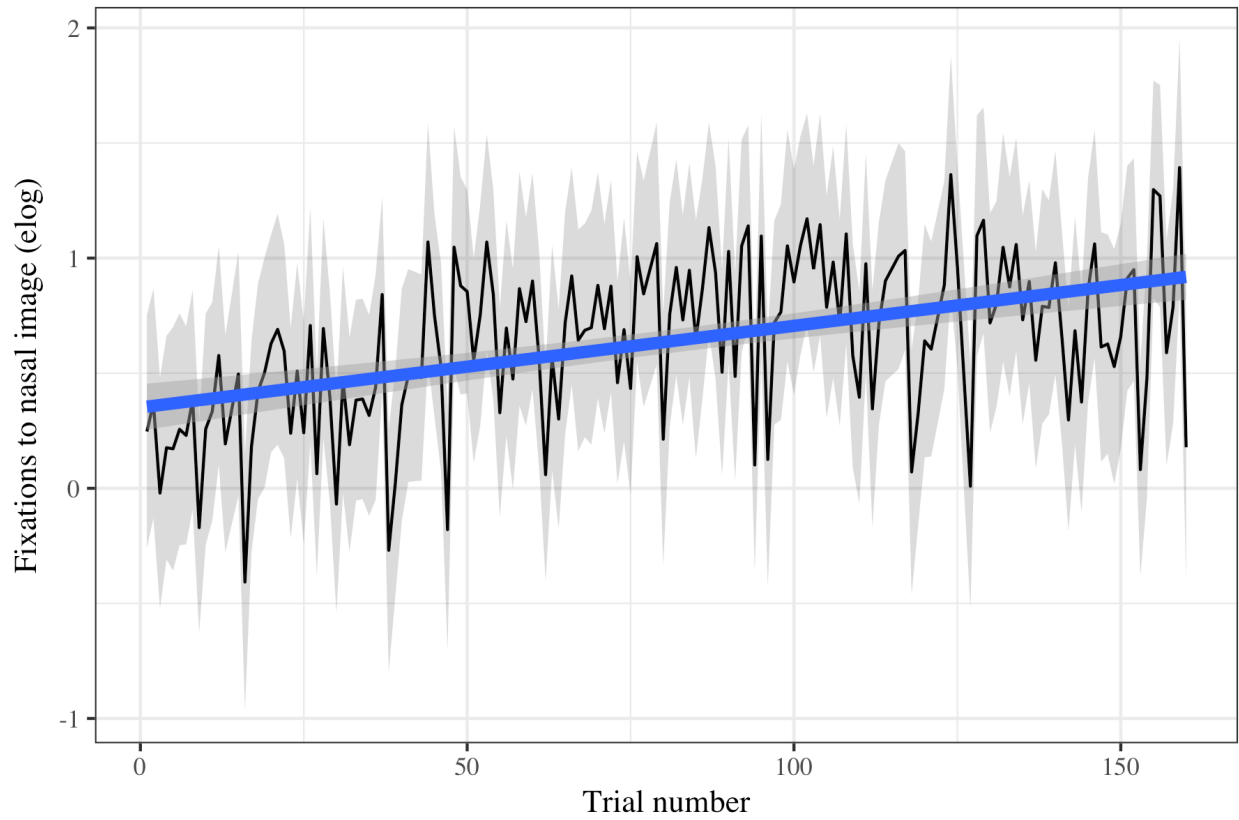


Figure A.29. Predicted data by TRIAL NUMBER – Exp. 2b elog %fix to $C\tilde{V}$, $C\tilde{V}$ choices

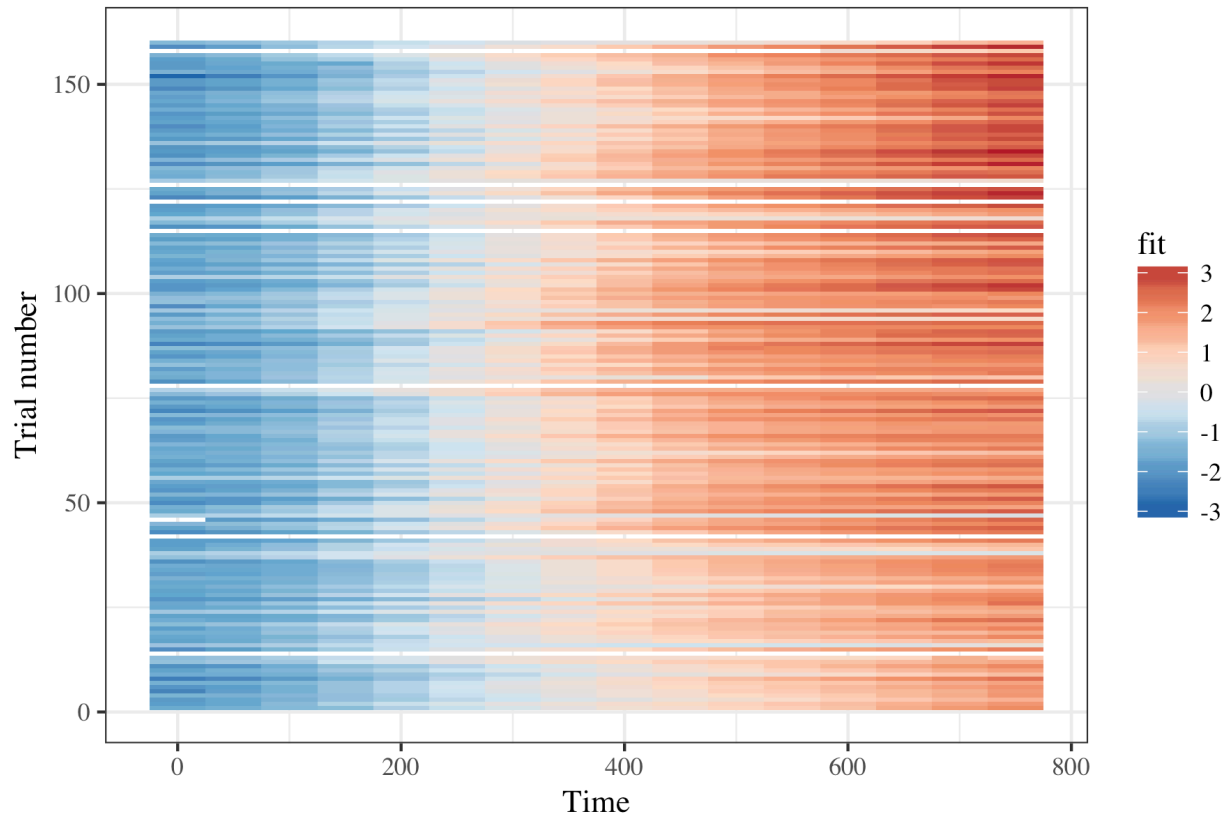


Figure A.30. Predicted data by TRIAL NUMBER through TIME – Exp. 2b elog %fix to $C\tilde{V}$, $C\tilde{V}$ choices

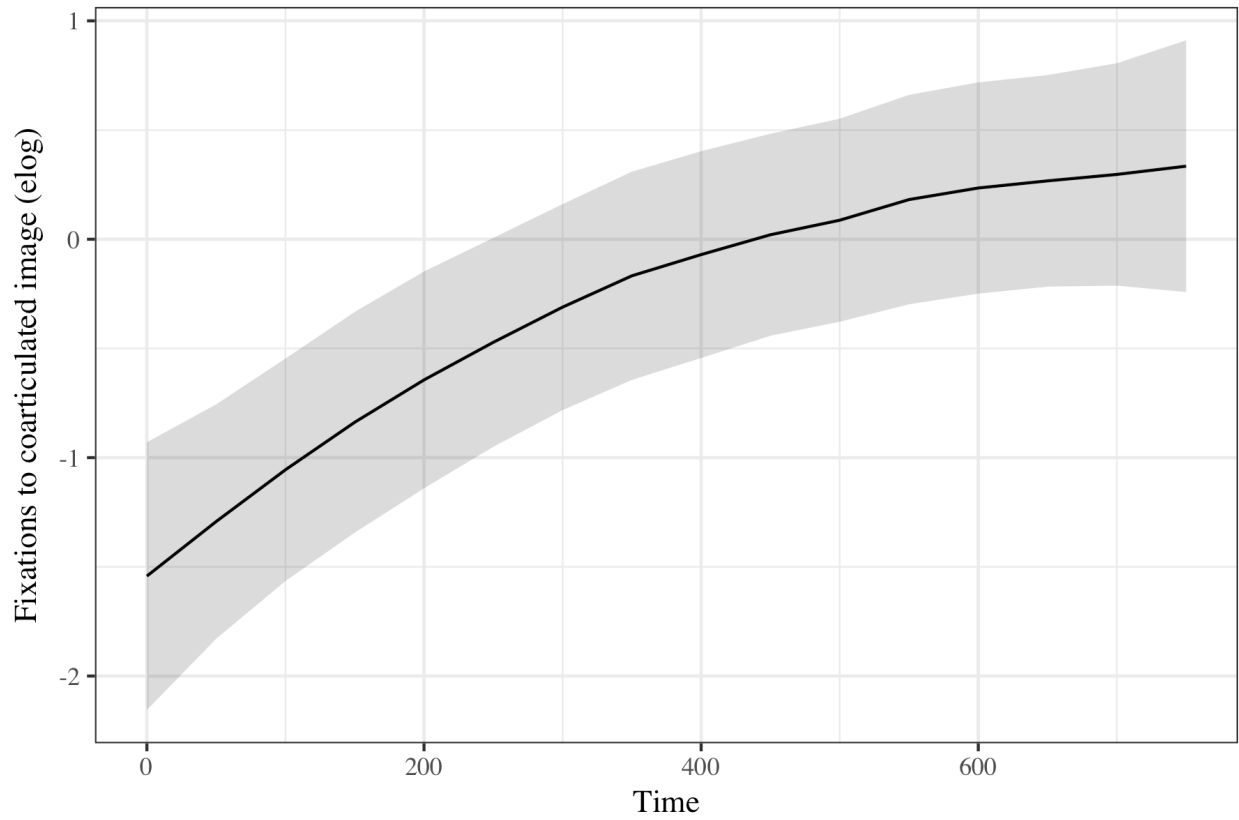


Figure A.31. Predicted data through TIME – Exp. 2b elog %fix to CVN, CVN choices

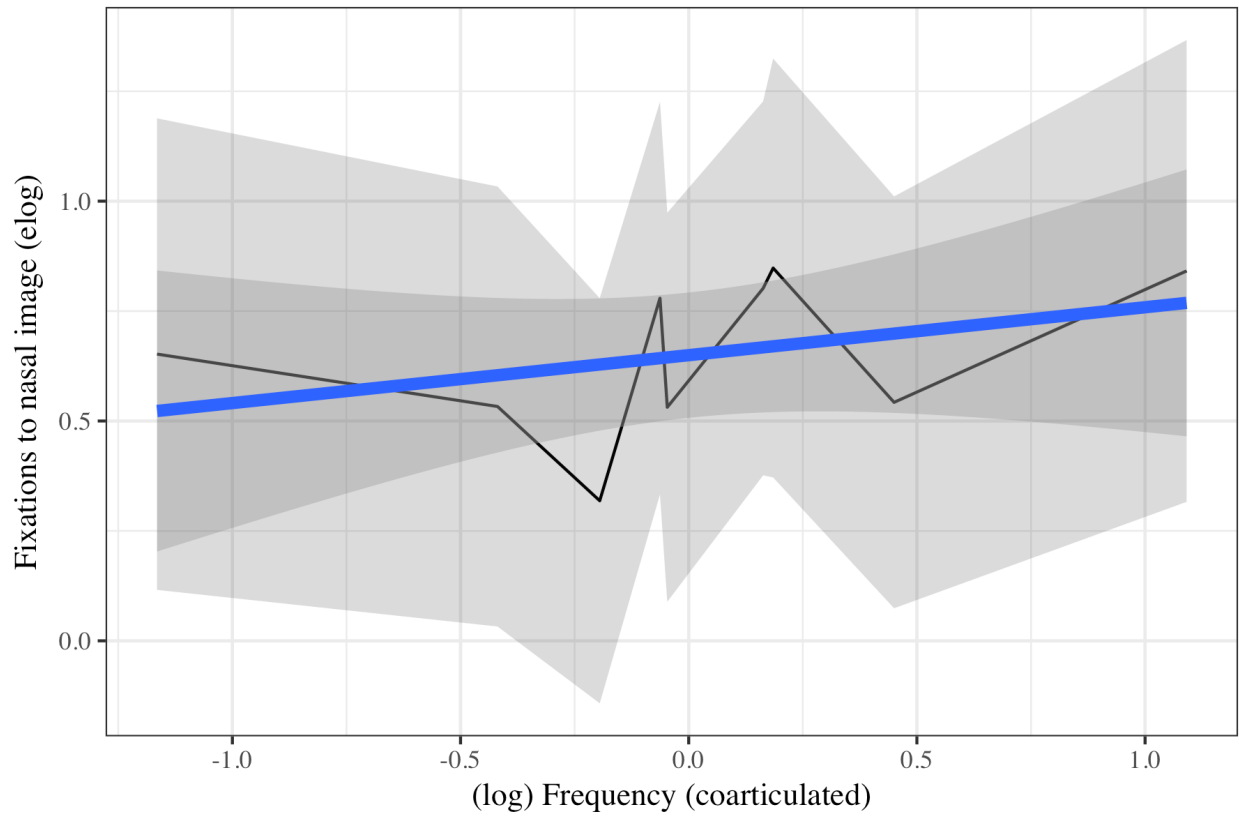


Figure A.32. Predicted data by FREQUENCY (COARTICULATED) – Exp. 2b elog %fix to CVN, CVN choices

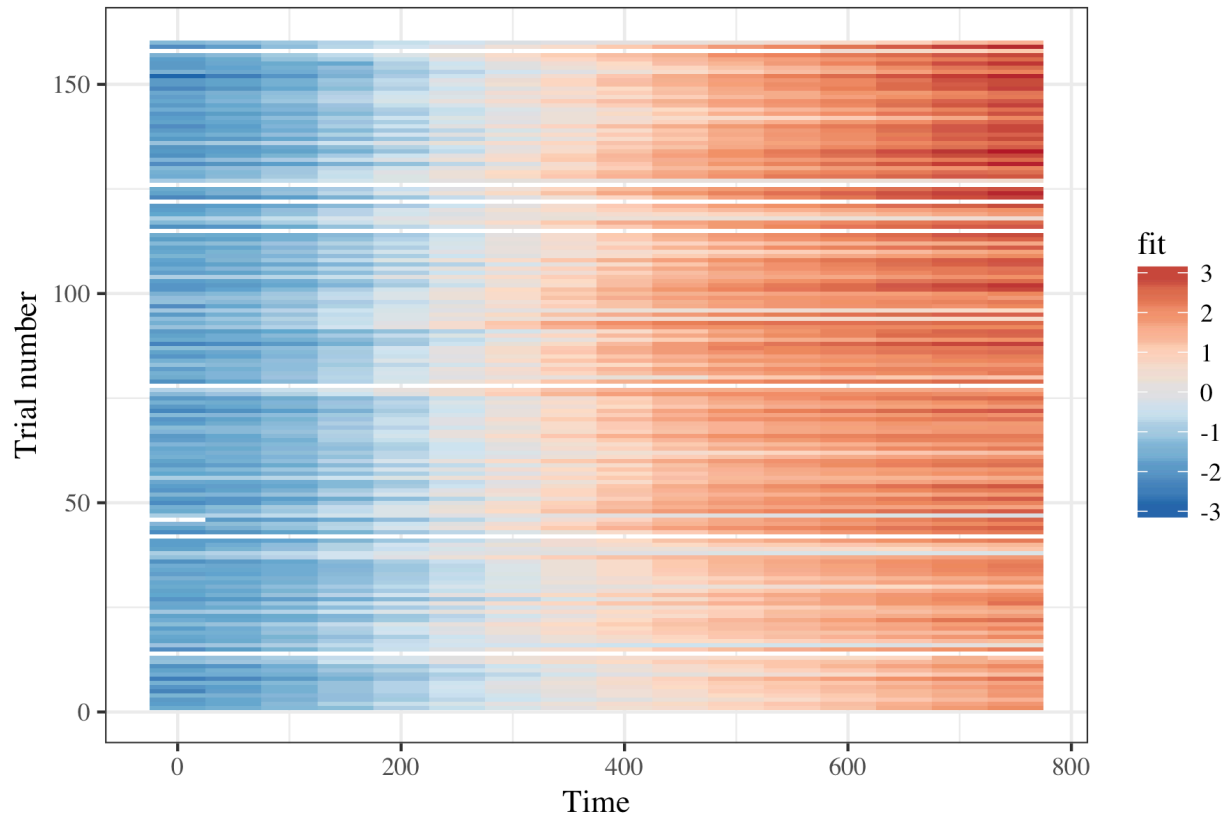


Figure A.33. Predicted data by TRIAL NUMBER through TIME – Exp. 2b elog %fix to CVN, CVN choices

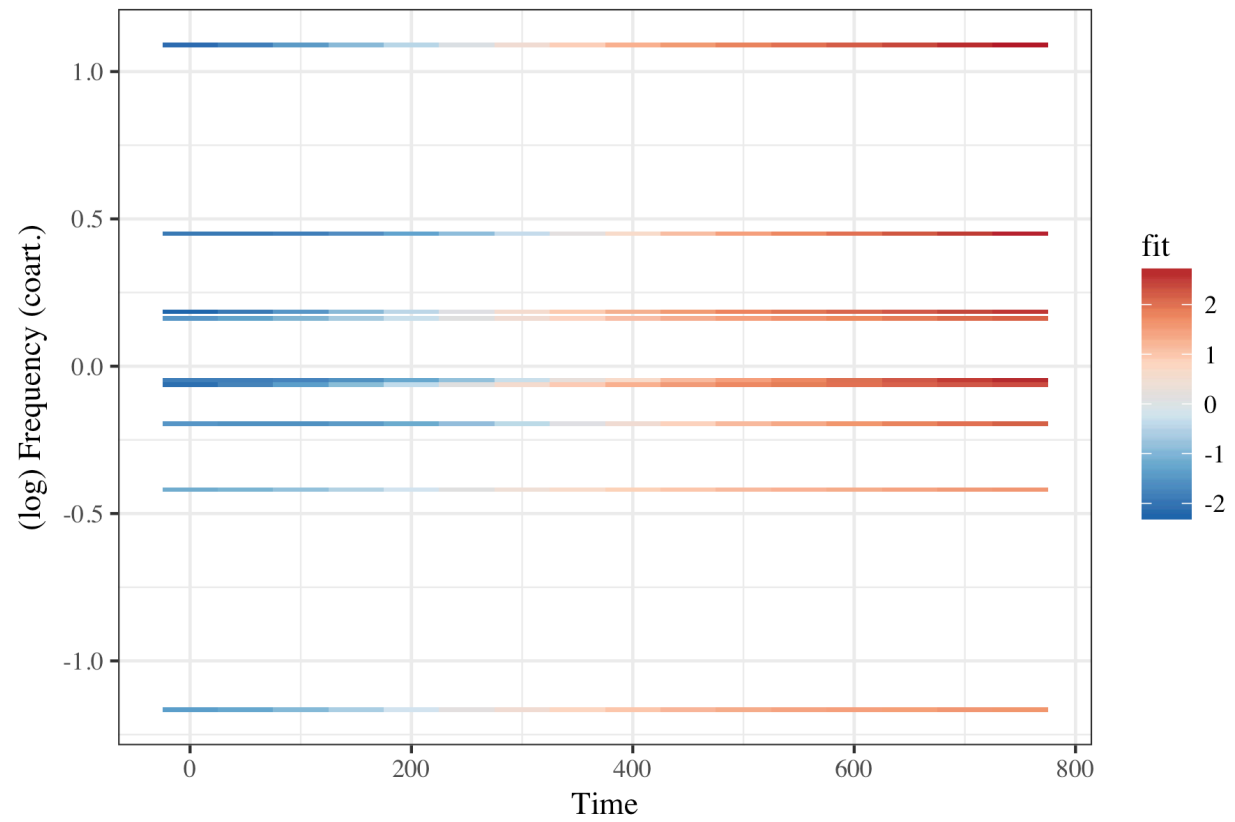


Figure A.34. Predicted data by FREQUENCY (COARTICULATED) through TIME – Exp. 2b elog %fix to CVN, CVN choices

APPENDIX 4

Table A.1

Results of the L2 familiarisation phase – number of TRÈS FAMILIER, ASSEZ FAMILIER, PEU FAMILIER, and PAS FAMILIER responses, collapsed for all L2 participants across both Experiments 2a and 2b (N=88)

CŨ word	Très familier	Assez familier	Peu familier	Pas familier	CVN word	Très familier	Assez familier	Peu familier	Pas familier
<i>bain</i>	67	12	6	3	<i>beigne</i>	52	15	11	10
<i>camp</i>	42	15	12	19	<i>canne</i>	54	22	6	6
<i>dent</i>	82	4	0	2	<i>dame</i>	67	11	6	4
<i>gant</i>	59	13	8	8	<i>gamme</i>	9	11	20	48
<i>pain</i>	84	2	1	1	<i>peigne</i>	25	24	15	24
<i>pont</i>	66	2	7	13	<i>pomme</i>	87	1	0	0
<i>rond</i>	59	19	9	1	<i>rhum</i>	48	24	10	6
<i>thon</i>	26	27	19	16	<i>tonne</i>	58	19	7	4
<i>vin</i>	75	8	0	5	<i>veine</i>	61	19	7	1