



HAL
open science

Perception and accommodation among French learners of English : an acoustic and electroglottographic study of creaky voice

Léa Burin

► **To cite this version:**

Léa Burin. Perception and accommodation among French learners of English : an acoustic and electroglottographic study of creaky voice. Linguistics. Université Paris Cité, 2022. English. NNT : 2022UNIP7095 . tel-04230698

HAL Id: tel-04230698

<https://theses.hal.science/tel-04230698v1>

Submitted on 6 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Perception and accommodation among French learners of English:

An acoustic and electroglottographic study of creaky
voice

THÈSE

présentée et soutenue publiquement le 25 novembre 2022

pour l’obtention du

Doctorat de l’Université Paris Cité

(mention Linguistique anglaise, spécialité phonétique/phonologie)

par

Léa Burin

sous la direction de Nicolas BALLIER

Composition du jury

Rapporteurs : Jane STUART-SMITH University of Glasgow
 Claire PILLOT-LOISEAU Université Sorbonne-Nouvelle

Examineurs : Richard WRIGHT University of Washington
 Emmanuel FERRAGNE Université Paris Cité

Acknowledgments

First and foremost, I am extremely grateful to Nicolas Ballier, my supervisor. You have guided and trusted me since the first days of my M1, seven years ago. Completing a PhD can be quite emotionally draining and doubts often surface. You have always been there for me, as a human and as a mentor, giving me freedom, but also pushing me when I needed to be pushed. These last few months in particular have been pretty intense and you have dedicated a lot of your time to me. Thank you again for your support, and for everything else. It has been a great journey (although a tough one).

Many thanks are due to Emmanuel Ferragne, who introduced me to experimental phonetics (to the fun stuff?). You dedicate a lot of your time promoting that branch of phonetics, and students like me can only be grateful. Thank you for all the help and advice you provided me along these years, and for answering emails so quickly (that has saved me a few anxiety-induced panic attacks).

I am also very grateful to Richard Wright, for his very warm welcome during my stay at the University of Washington. You took me under your wing, introduced me to members of the lab and to how life is back there. More importantly, you provided considerable help with my experimental design. It has been a great pleasure to work with someone that passionate and fascinating. Thank you for your devotion and humanism, my PhD thesis would not have been what it is if you had not been there.

Finally, many thanks to the reviewers of this thesis, Claire Pillot-Loiseau and Jane Stuart-Smith, for their patience.

I would also like to thank

All the linguistic students, professors, researchers I met during my degree and my years as a PhD student. I would particularly like to thank Alain Diana, who introduced me to the study of speech sounds during the first year of my BA. If I specialised in phonetics and phonology, it is thanks to you.

All the people who contributed to my research in one way or another. Thanks to the ED 622 and CLILLAC-ARP to have placed your trust in me for a PhD grant. Thanks to the Labex-EFL for their grant that allowed me to go to the United States to collect part of my data, and to work alongside Richard Wright. Also many thanks to Laura Panfili, Marc Garellek, and Yen Shue for their help with VoiceSauce; Patricia Keating for answering all of my questions on creaky voice, and Marina Oganyan for her help with PsychoPy.

All who volunteered to take part in my experiment, and those I have tortured with the electroglottograph, either at UW or at Université Paris Cité. You have been amazing guinea pigs.

My family, my dad and Popointu and, more importantly, my brother and my sister, *mes trésors*, my biggest supporters. You are the most extraordinary and unique people on this planet, and I do not know what I would do without you. I guess it was not too bad of an idea to follow your steps and study English?

My very best friend, Nab. You have perfected the art of synthesising my research to other people with years (better than I have). Thank you for everything you have done for me, and for being the wonderful person that you are. I am extremely proud to count you as a friend.

My girl gang, Nab, Bayou, Paupau, Loulou, Chlochlo, and Momo. I cannot believe it has been 13 years since I met you all. May our adventures and friendship last forever!

My dearest and closest friends, Hajare, Robin, Pess, Jonas, Max, Louis, Mathilde, and to all those who have been there at some point. Your favourite question now has an answer: “I’m done with the PhD”.

Babs, I would not have met you if I had not done that PhD. Thanks for always calming and reassuring me, and for letting me vent all of my frustration at you. You have been one massive support. You are the best.

Last but not least, *le bureau 825*. Thanks to everyone I shared the office with during those years. More particularly, I would like to thank Léo, Coline, Alessandro, and Cameron (/ˈkæməɾən/). These past few years would not have been the same without you, and it made that whole journey way easier (and way more fun).

Léo, we did it.

*To my mother.
'Le plus fort du monde entier'.*

Contents

| | |
|--|-------------|
| Acknowledgments | i |
| List of Figures | xi |
| List of Tables | xvii |
| Introduction | 3 |
| CHAPTER I: Phonation | 11 |
| 1 Physiology and phonation types | 12 |
| 1.1 Anatomy and physiology of the larynx | 12 |
| 1.1.1 General description | 12 |
| 1.1.2 Cartilaginous structure | 12 |
| 1.1.3 Muscular structure | 13 |
| 1.1.4 Differences between men and women | 15 |
| 1.2 Physiological correlates of phonation types | 16 |
| 1.3 Phonation types | 17 |
| 2 Acoustic measures of phonation types | 22 |
| 3 The different functions of non-modal phonation types | 26 |
| 3.1 Contrastive | 26 |
| 3.2 Allophonic | 27 |
| 3.3 Prosodic | 27 |
| 3.3.1 Phrase boundary | 28 |
| 3.3.2 Prominence | 28 |
| 3.3.3 Phrase boundary and prominence | 29 |
| 3.4 Paralinguistic | 29 |
| 3.5 Sociolinguistic | 31 |
| 3.5.1 Voice quality and language status | 31 |
| 3.5.2 Voice quality as an index of group membership | 32 |
| 3.5.3 Voice quality and role identity | 33 |

| | | |
|---|---|-----------|
| 4 | Conclusion | 34 |
| CHAPTER II: Phonetic convergence | | 35 |
| 1 | Measuring accommodation | 37 |
| 1.1 | Levels of measurement | 37 |
| 1.2 | Popular methodological practices | 37 |
| 2 | The psycholinguistic approach on phonetic accommodation | 38 |
| 2.1 | Convergence as an automatic process | 38 |
| 2.2 | The effect of language status | 41 |
| 2.3 | Perception of L2 sounds | 43 |
| 2.3.1 | Models of L2 perception and acquisition | 43 |
| 2.3.2 | The effect of proficiency | 44 |
| 2.3.3 | The effect of 'phonetic talent' | 45 |
| 3 | The sociolinguistic approach on phonetic accommodation | 46 |
| 3.1 | The effect of gender | 46 |
| 3.2 | The effect of interactional dominance | 47 |
| 3.3 | The effect of cultural affiliation | 48 |
| 3.4 | The effect of social preference | 49 |
| 4 | Creaky voice accommodation | 50 |
| 5 | Conclusion | 51 |
| CHAPTER III: Materials and Methods | | 53 |
| 1 | Recording procedures | 55 |
| 1.1 | Electroglottography | 55 |
| 1.2 | The electroglottograph | 56 |
| 1.3 | The EGG signal | 57 |
| 2 | Speech stimuli | 60 |
| 2.1 | Model speakers | 61 |
| 2.2 | Instructions | 61 |
| 2.3 | Recorded sentences | 61 |
| 2.4 | Segmentation | 63 |
| 2.5 | Selected stimuli | 64 |
| 3 | Participants | 65 |
| 4 | Procedure | 66 |
| 4.1 | Language proficiency test | 66 |
| 4.2 | Speech tasks | 68 |
| 4.2.1 | Reading task | 68 |

| | | | |
|---|-------|--|----|
| | 4.2.2 | Imitation task | 68 |
| | 4.2.3 | Judgment rating task | 69 |
| 5 | | Hypotheses | 70 |
| | 5.1 | Accommodation | 70 |
| | 5.2 | Evaluation | 71 |
| 6 | | Measurements | 72 |
| | 6.1 | Acoustic measures | 73 |
| | | 6.1.1 Duration | 73 |
| | | 6.1.2 Fundamental frequency | 74 |
| | | 6.1.3 Spectral tilt | 75 |
| | | 6.1.4 Cepstral Peak Prominence (CPP) | 77 |
| | | 6.1.5 Summary of the acoustic measurements | 78 |
| | 6.2 | Electroglottographic measures | 79 |
| | | 6.2.1 Closed quotient | 79 |
| | | 6.2.2 PIC (Peak Increase in Contact) | 79 |
| 7 | | Software | 79 |
| | 7.1 | EggWorks | 79 |
| | 7.2 | VoiceSauce | 82 |
| 8 | | Conclusion | 84 |

CHAPTER IV: A comparative analysis of creaky voice across language and gender **85**

| | | | |
|---|------|---|-----|
| 1 | | Data processing | 87 |
| 2 | | Data visualization | 87 |
| 3 | | Model selection | 88 |
| 4 | | Analysis of acoustic measures | 89 |
| | 4.1 | Duration | 89 |
| | 4.2 | Fundamental frequency | 94 |
| | 4.3 | H1* | 97 |
| | 4.4 | H2* | 101 |
| | 4.5 | H1*-H2* | 104 |
| | 4.6 | H1*-A1* | 107 |
| | 4.7 | H1*-A2* | 110 |
| | 4.8 | H1*-A3* | 114 |
| | 4.9 | H2*-H4* | 116 |
| | 4.10 | CPP | 119 |
| 5 | | Analysis of EGG measures | 125 |

| | | |
|--|---|------------|
| 5.0.1 | Closed quotient | 125 |
| 5.1 | PIC | 127 |
| 6 | Principal component analyses | 132 |
| 6.1 | Data normalisation | 132 |
| 6.2 | Preliminary analysis | 132 |
| 6.3 | Correlation analysis | 133 |
| 6.4 | Variance analysis | 135 |
| 6.5 | Cross-gender differences within-language analysis | 137 |
| 7 | Conclusion | 138 |
| CHAPTER V: Accommodation and perception of creaky voice | | 141 |
| 1 | Reading task | 143 |
| 1.1 | Overall observations | 143 |
| 1.2 | Between-speaker variability | 144 |
| 1.3 | The effect of vowel type | 146 |
| 2 | Imitation task | 149 |
| 2.1 | Difference-in-Distance | 149 |
| 2.2 | Data visualization | 150 |
| 2.3 | Model selection | 150 |
| 2.4 | Analysis of acoustic measures | 150 |
| 2.4.1 | Duration DID score | 150 |
| 2.4.2 | Fundamental frequency DID score | 154 |
| 2.4.3 | H1* DID score | 158 |
| 2.4.4 | H2* DID score | 161 |
| 2.4.5 | H1*-H2* DID score | 164 |
| 2.4.6 | H1*-A1* DID score | 168 |
| 2.4.7 | H1*-A2* DID score | 170 |
| 2.4.8 | H1*-A3* DID score | 173 |
| 2.4.9 | H2*-H4* DID score | 176 |
| 2.4.10 | CPP DID score | 180 |
| 2.5 | Analysis of EGG measures | 183 |
| 2.5.1 | Closed quotient DID score | 183 |
| 2.5.2 | PIC DID score | 187 |
| 3 | Rating task | 194 |
| 3.1 | Overall observations | 194 |
| 3.2 | Language effect | 196 |
| 3.2.1 | American English | 196 |

| | | |
|--|--|------------|
| 3.2.2 | French | 197 |
| 3.2.3 | American English vs. French | 198 |
| 3.3 | Gender effect within-language | 199 |
| 3.3.1 | American English male | 199 |
| 3.3.2 | American English female | 201 |
| 3.3.3 | French male | 202 |
| 3.3.4 | French female | 203 |
| 3.3.5 | Male vs. Female American English | 203 |
| 3.3.6 | Male vs. Female French | 204 |
| 3.4 | Model speaker effect | 205 |
| 3.4.1 | Pleasant | 206 |
| 3.4.2 | Attractive | 208 |
| 3.4.3 | Powerful | 210 |
| 3.4.4 | Educated | 212 |
| 4 | Conclusion | 215 |
| CHAPTER VI: General conclusion and discussion | | 219 |
| 1 | Creaky voice | 219 |
| 2 | The present study | 221 |
| 3 | Summary of the results and discussion | 223 |
| 3.1 | Comparative analysis of creaky voice across language and gender | 223 |
| 3.2 | A comparison across tasks | 226 |
| 3.2.1 | Reading task | 226 |
| 3.2.2 | Repetition/Imitation task | 227 |
| 3.2.3 | Judgment rating task | 230 |
| 4 | Methodological issues | 232 |
| 5 | Future research | 233 |
| 5.1 | With current data | 233 |
| 5.2 | With additional data | 235 |
| 5.3 | With new data | 236 |
| Bibliography | | 239 |
| Appendix A Stimuli | | 263 |
| A.1 | List of the English sentences presented to the American model speakers | 264 |
| A.2 | List of the French sentences presented to the French model speakers | 265 |
| A.3 | Phonological transcription of English and French sentence-final words | 268 |
| A.4 | Key to phonetic/SAMPA symbols for English and French | 269 |

| | |
|--|------------|
| A.5 List of the stimuli selected for each model talker | 274 |
| Appendix B Analysis | 277 |
| B.1 Convergence and divergence effects for each measure depending on model speaker | 277 |
| B.2 Linear mixed-effects models | 280 |
| Appendix C Metadata | 287 |
| C.1 Model speakers | 288 |
| C.2 Participants | 294 |
| Appendix D IRB Requirements | 299 |
| Résumé | 305 |
| Abstract | 306 |

List of Figures

| | | |
|----|--|-----|
| 1 | Cartilaginous structure of the larynx (from: Jones & Barnes, 2019). | 12 |
| 2 | Intrinsic muscles of the larynx (from: The Anatomical Chart Series, 1993). | 14 |
| 3 | Laryngeal parameters (from: Wright <i>et al.</i> , 2019). | 16 |
| 4 | Phonation types and their relationship to glottal aperture (from: Wright <i>et al.</i> , 2019). | 17 |
| 4 | The EGG setup. | 56 |
| 5 | Divisions of the vibratory or vocal-fold contact cycle (from: Baken & Orlikoff, 2000). | 58 |
| 6 | Figure depicting a glottal cycle as observed on an schematic and real EGG signal and its derivative (DEGG). (1)-(3): closing phase, (3)-(4): closed phase, (4)-(6): opening phase, (6)-(1): open phase. A schematic representation of the corresponding glottal flow (ODG) has been added as a comparison (from: Henrich, 2001). | 60 |
| 7 | Example displaying file segmentation for a model speaker. | 63 |
| 8 | Screenshot of what participants would see during the judgment rating task. | 70 |
| 9 | EggWorks parameters as found on the software page. | 79 |
| 10 | Measures output from EggWorks (Tehrani, 2009) showing CQ, CQ_H, CQ_PM and CQ_HT. This figure was extracted from Keating <i>et al.</i> , 2012. | 82 |
| 11 | The derivative of the signal on which is indicated the peak decrease in contact (PDC) and the peak increase in contact (PIC), which corresponds to the peak_Vel measure in EggWorks. This figure was extracted from Keating <i>et al.</i> , 2012. | 82 |
| 12 | VoiceSauce settings interface. Parameters framed in red were modified. Default values were used for all other parameters. | 84 |
| 13 | Violin plot representing the distribution of creaky vowels' duration across language and gender. | 90 |
| 14 | Effect of language and gender on creaky vowels' duration. The shaded areas represent the standard deviation for each group. | 92 |
| 15 | Violin plot representing the distribution of creaky vowels' duration across model speakers. | 93 |
| 16 | Effect of vowel type on creaky vowels' duration. The shaded area represents the standard deviation. | 94 |
| 17 | Violin plot representing the distribution of creaky vowels' f_0 values across language and gender. | 95 |
| 18 | Effect of gender, on the left, and language, on the right, on creaky vowels's f_0 | 97 |
| 19 | Violin plot representing the distribution of creaky vowels' H1* values across language and gender. | 98 |
| 20 | Violin plot representing the distribution of creaky vowels' H1* values across model speaker. | 99 |
| 21 | Violin plot representing the distribution of creaky vowels' H1* values across vowels. | 100 |

| | | |
|----|--|-----|
| 22 | Violin plot representing the distribution of creaky vowels' H2* values across language and gender. | 101 |
| 23 | Violin plot representing the distribution of creaky vowels' H2* values across model speakers. | 104 |
| 24 | Effect of vowel type on creaky vowels' H2*. The shaded area represents the standard deviation. | 105 |
| 25 | Violin plot representing the distribution of creaky vowels' H1*-H2* values across language and gender. | 105 |
| 26 | Violin plot representing the distribution of creaky vowels' H1*-A1* values across language and gender. | 107 |
| 27 | Violin plot representing the distribution of creaky vowels' H1*-A1* across model speakers. | 109 |
| 28 | Violin plot representing the distribution of creaky vowels' H1*-A1* across vowels. | 110 |
| 29 | Violin plot representing the distribution of creaky vowels' H1*-A2* values across language and gender. | 111 |
| 30 | Violin plot representing the distribution of creaky vowels' H1*-A2* across model speakers. | 113 |
| 31 | Violin plot representing the distribution of creaky vowels' H1*-A2* across vowels. | 114 |
| 32 | Violin plot representing the distribution of creaky vowels' H1*-A3* values across language and gender. | 114 |
| 33 | Violin plot representing the distribution of creaky vowels' H1*-A3* values across model speakers. | 117 |
| 34 | Violin plot representing the distribution of H1*-A3* values across vowels. | 117 |
| 35 | Violin plot representing the distribution of creaky vowels' H2*-H4* values across language and gender. | 118 |
| 36 | Effect of gender on creaky vowels' H2*-H4*. The shaded area represents the standard deviation. | 120 |
| 37 | Violin plot representing the distribution of creaky vowels' CPP values across language and gender. | 120 |
| 38 | Effect of language on creaky vowels' CPP. The shaded area represents the standard deviation. | 123 |
| 39 | Violin plot representing the distribution of CPP values across model speakers. | 123 |
| 40 | Effect of vowel type on creaky vowels' CPP. The shaded area represents the standard deviation. | 124 |
| 41 | Violin plot representing the distribution of creaky vowels' CQ values across language and gender. | 125 |
| 42 | Violin plot representing the distribution of creaky vowels' CQ values across model speakers. | 127 |
| 43 | Violin plot representing the distribution of creaky vowels' PIC values across language and gender. | 128 |
| 44 | Interaction effect between language and gender on creaky vowels' PIC. | 130 |
| 45 | Violin plot representing the distribution of creaky vowels' PIC values across model speakers. | 131 |
| 46 | Scree plot representing the number of dimensions and their corresponding percentage of explained variances resulting from the PCA analyses of French and American English. | 132 |
| 47 | Correlation circles for American English and French. | 133 |
| 48 | Contribution of variables for PC1 and PC2 in both French and American English. | 136 |
| 49 | Variation across language and gender in both PC1 and PC2. | 137 |
| 50 | Mean proportion of creaky vowels in the reading task depending on language. Error bars represent the standard error of the mean. | 143 |
| 51 | Mean proportion of word-final creak in the reading task depending on language and speaker. | 144 |

| | | |
|----|---|-----|
| 52 | Correlation and regression plot showing the proportion of word-final creaky vowels (%) in English L2 as compared to that in French L1. | 145 |
| 53 | Creaky and non-creaky realisations of vowels according to vowel type in the reading task. | 146 |
| 54 | Mean proportion of low creaky vowels, on the left panel, and of high creaky vowels, on the right panel, across language. Error bars represent the standard error of the mean. | 147 |
| 55 | Creaky and non-creaky realisations of vowels in the reading task. | 148 |
| 56 | Bar plot representing the mean duration DID score by language and gender. Error bars represent the standard error of the mean. | 151 |
| 57 | Effect of language and gender on duration DID scores. The shaded areas represent the standard deviation. | 153 |
| 58 | Effect of model speaker on duration DID scores. Error bars represent the standard error of the mean. | 153 |
| 59 | Bar plot representing the mean f_0 DID score by language and gender. Error bars represent the standard error of the mean. | 154 |
| 60 | Effect of language on f_0 DID scores. The shaded area represents the standard deviation of the mean. | 156 |
| 61 | Effect of language and gender on f_0 DID scores. The shaded areas represent the standard deviation of the mean. | 156 |
| 62 | Effect of model speaker on f_0 DID scores. Error bars represent the standard error of the mean. | 157 |
| 63 | Bar plot representing the mean H1* DID score by language and gender. Error bars represent the standard error of the mean. | 158 |
| 64 | Effect of language on H1* DID scores. The shaded area represents the standard deviation. | 160 |
| 65 | Effect of language and gender on H1* DID scores. The shaded areas represent the standard deviation. | 160 |
| 66 | Effect of model speaker on H1* DID scores. Error bars represent the standard error of the mean. | 161 |
| 67 | Bar plot representing the mean H2* DID score by language and gender. Error bars represent the standard error of the mean. | 162 |
| 68 | Effect of language on H2* DID scores. The shaded area represents the standard deviation. | 163 |
| 69 | Effect of language and gender on H2* DID scores. The shaded area represents the standard deviation. | 164 |
| 70 | Effect of model speaker on H2* DID scores. Error bars represent the standard deviation of the mean. | 165 |
| 71 | Bar plot representing the mean H1*-H2* DID score by language and gender. Error bars represent the standard error of the mean. | 166 |
| 72 | Effect of language on H1*-H2* DID scores. The shaded area represents the standard deviation. | 167 |
| 73 | Effect of model speaker on H1*-H2* DID scores. Error bars represent the standard deviation of the mean. | 168 |
| 74 | Bar plot representing the mean H1*-A1* DID score by language and gender. Error bars represent the standard error of the mean. | 169 |
| 75 | Bar plot representing the mean H1*-A2* DID score by language and gender. Error bars represent the standard error of the mean. | 171 |
| 76 | Effect of language and gender on H1*-A2* DID scores. The shaded areas represent the standard deviation. | 172 |

| | | |
|----|--|-----|
| 77 | Effect of model speaker on H1*-A2* DID scores. Error bars represent the standard deviation of the mean. | 173 |
| 78 | Bar plot representing the mean H1*-A3* DID score by language and gender. Error bars represent the standard error of the mean. | 174 |
| 79 | Effect of language on H1*-A3* DID scores. The shaded area represents the standard deviation. | 175 |
| 80 | Effect of language and gender on H1*-A3* DID scores. The shaded areas represent the standard deviation. | 176 |
| 81 | Effect of model speaker on H1*-A3* DID scores. Error bars represent the standard deviation of the mean. | 177 |
| 82 | Bar plot representing the mean H2*-H4* DID score by language and gender. Error bars represent the standard error of the mean. | 178 |
| 83 | Effect of model speaker on H2*-H4* DID scores. Error bars represent the standard deviation of the mean. | 179 |
| 84 | Bar plot representing the mean CPP DID score by language and gender. Error bars represent the standard error of the mean. | 180 |
| 85 | Effect of language on CPP* DID scores. The shaded area represents the standard deviation. | 182 |
| 86 | Effect of language and gender on CPP DID scores. The shaded areas represent the standard deviation. | 182 |
| 87 | Effect of model speaker on CPP DID scores. Error bars represent the standard deviation of the mean. | 183 |
| 88 | Bar plot representing the mean CQ DID score by language and gender. Error bars represent the standard error of the mean. | 184 |
| 89 | Effect of language and gender on CQ DID scores. The shaded areas represent the standard deviation. | 186 |
| 90 | Effect of model speaker on CQ DID scores. Error bars represent the standard deviation of the mean. | 186 |
| 91 | Bar plot representing the mean PIC DID score by language and gender. Error bars represent the standard error of the mean. | 187 |
| 92 | Effect of language on PIC DID scores. The shaded area represents the standard deviation. | 189 |
| 93 | Effect of language and gender on PIC DID scores. The shaded areas represent the standard deviation. | 189 |
| 94 | Effect of model speaker on PIC DID scores. Error bars represent the standard deviation of the mean. | 190 |
| 95 | Voice quality rating across the whole dataset. Creaky voice is represented by the pink line, non-creaky voice by the green one. Shaded areas represent the standard deviation of the mean. | 195 |
| 96 | Voice quality rating across language. Ratings of American English voices are displayed on the left panel, those of French on the right panel. Creaky voice is represented by the pink line, non-creaky voice by the green one. Shaded areas represent the standard deviation of the mean. | 196 |
| 97 | Voice quality rating across language and gender. Ratings of American English voices are displayed on the left panel, those of French on the right. Ratings of male voices are displayed on the top panels, those of female on the bottom ones. Creaky voice is represented by the pink line, non-creaky voice by the green one. Shaded areas represent the standard deviation of the mean. | 200 |

| | | |
|-----|--|-----|
| 98 | Voice quality rating across model speakers. One colour corresponds to one speaker. The continuous represents creaky voice and the dashed line non-creaky voice. Shaded areas represent the standard deviation of the mean. | 206 |
| 99 | Probability density function of PLEASANT for the different model speakers. Ratings of creaky voice are illustrated on the left panel, those of non-creaky voice on the right panel. | 207 |
| 100 | Probability density function of ATTRACTIVE for the different model speakers. Ratings of creaky voice are illustrated on the left panel, those of non-creaky voice on the right panel. | 209 |
| 101 | Probability density function of POWERFUL for the different model speakers. Ratings of creaky voice are illustrated on the left panel, those of non-creaky voice on the right panel. | 211 |
| 102 | Probability density function of EDUCATED for the different model speakers. Ratings of creaky voice are illustrated on the left panel, those of non-creaky voice on the right panel. | 213 |

List of Tables

| | | |
|----|--|-----|
| 1 | Results obtained to all tests by each selected participant. For more visibility, B2 level is indicated in red, C1 level in blue, and C2 level in green | 67 |
| 2 | Hypotheses and predictions. | 72 |
| 3 | Summary of the acoustic measurements extracted with VoiceSauce (from: Shue <i>et al.</i> , 2011). (*) indicates that the harmonic and spectral amplitudes were corrected for formant frequencies and bandwidths. | 78 |
| 4 | EggWorks measurements output as taken directly from the software webpage. Measurements marked with an asterisk are those used in the study. | 81 |
| 5 | Mean creaky vowels' duration (in ms) across language and gender. Standard deviation is indicated in parenthesis. | 90 |
| 6 | Significant and non-significant effects for duration (Marginal $R^2 = 0.423$ /Conditional $R^2 = 0.707$). | 91 |
| 7 | Mean f_0 values (in Hz) across language and gender. Standard deviation is indicated in parenthesis. | 95 |
| 8 | Significant and non-significant effects for f_0 (Marginal $R^2 = 0.52$ /Conditional $R^2 = 0.539$). | 96 |
| 9 | Mean H1* values (in dB) across language and gender. Standard deviation is indicated in parenthesis. | 98 |
| 10 | Significant and non-significant effects for H1* (Marginal $R^2 = 0.186$ /Conditional $R^2 = 0.795$). | 99 |
| 11 | Mean H2* values (in dB) across language and gender. Standard deviation is indicated in parenthesis. | 102 |
| 12 | Significant and non-significant effects for H2* (Marginal $R^2 = 0.106$ /Conditional $R^2 = 0.6$). | 103 |
| 13 | Mean H1*-H2* values (in dB) across language and gender. Standard deviation is indicated in parenthesis. | 106 |
| 14 | Mean H1*-A1* values (in dB) across language and gender. Standard deviation is indicated in parenthesis. | 108 |
| 15 | Significant and non-significant effects for H1*-A1* (Marginal $R^2 = 0.071$ /Conditional $R^2 = 0.556$). | 109 |
| 16 | Mean H1*-A2* values (in dB) across language and gender. Standard deviation is indicated in parenthesis. | 111 |
| 17 | Significant and non-significant effects for H1*-A2* (Marginal $R^2 = 0.155$ /Conditional $R^2 = 0.476$). | 112 |
| 18 | Mean H1*-A3* values (in dB) across language and gender. Standard deviation is indicated in parenthesis. | 115 |

| | | |
|----|--|-----|
| 19 | Significant and non-significant effects for H1*-A3* (Marginal R ² = 0.042/Conditional R ² = 0.408). | 116 |
| 20 | Mean H2*-H4* values (in dB) across language and gender. Standard deviation is indicated in parenthesis. | 118 |
| 21 | Significant and non-significant effects for H2*-H4* (Marginal R ² = 0.157/Conditional R ² = 0.235). | 119 |
| 22 | Mean CPP values (in dB) across language and gender. Standard deviation is indicated in parenthesis. | 121 |
| 23 | Significant and non-significant effects for CPP (Marginal R ² = 0.265/Conditional R ² = 0.544). | 122 |
| 24 | Mean CQ values across language and gender. Standard deviation is indicated in parenthesis | 126 |
| 25 | Significant and non-significant effects for CQ (Marginal R ² = 0.138/Conditional R ² = 0.498). | 126 |
| 26 | Mean PIC values across language and gender. Standard deviation is indicated in parenthesis. FM is for French males, FF for French females, AM for American males, and AF for American females | 128 |
| 27 | Significant and non-significant effects for PIC (Marginal R ² = 0.420/Conditional R ² = 0.716). | 129 |
| 28 | Synthesis of the models found for each variable and their significant effects. | 140 |
| 29 | Mean vowel duration difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively). | 152 |
| 30 | Mean f_0 difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively). | 155 |
| 31 | Mean H1* difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively). | 159 |
| 32 | Mean H2* difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively). | 162 |
| 33 | Mean H1*-H2* difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively). | 166 |
| 34 | Mean H1*-A1* difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively). | 169 |
| 35 | Mean H1*-A2* difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively). | 171 |
| 36 | Mean H1*-A3* difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively). | 174 |
| 37 | Mean H2*-H4* difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively). | 177 |

| | | |
|------|--|-----|
| 38 | Mean CPP difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively). | 181 |
| 39 | Mean CQ difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively). | 184 |
| 40 | Mean PIC difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively). | 188 |
| 41 | Synthesis of the significant patterns of accommodation observed on acoustic measures in our dataset. Overall significant patterns are reported in the left column, those across language in the middle column, those resulting from the interaction between Language and Gender in the right column. An asterisk has been added whenever the pattern was the same across groups to indicate what group received the most convergence/divergence. Blank spaces mean no significant patterns of accommodation. | 192 |
| 42 | Synthesis of the significant patterns of accommodation observed on EGG measures in our dataset. Overall significant patterns are reported in the left column, those across language in the middle column, those resulting from the interaction between Language and Gender in the right column. An asterisk has been added whenever the pattern was the same across groups to indicate what group received the most convergence/divergence. Blank spaces mean no significant patterns of accommodation. | 193 |
| A.1 | Phonological transcription of sentence-final words. | 268 |
| A.2 | American English consonants. | 269 |
| A.3 | American English vowels. | 270 |
| A.4 | French consonants. | 271 |
| A.5 | French vowels. | 272 |
| A.6 | List of the stimuli selected for each model speaker. | 275 |
| B.1 | Convergence and divergence effects for each measure depending on model speaker. C is indicated when convergence was observed, D when divergence was observed. | 278 |
| B.2 | Best model found for duration DID score (Conditional $R^2 = 0.072$). | 280 |
| B.3 | Best model found for f_0 DID score (Conditional $R^2 = 0.039$). | 281 |
| B.4 | Best model found for H1* DID score (Conditional $R^2 = 0.228$). | 281 |
| B.5 | Best model found for H2* DID score (Conditional $R^2 = 0.179$). | 282 |
| B.6 | Best model found for H1*-H2* DID score (Conditional $R^2 = 0.018$). | 282 |
| B.7 | Best model found for H1*-A1* DID score (Conditional $R^2 = 0.015$). | 283 |
| B.8 | Best model found for H1*-A2* DID score (Conditional $R^2 = 0.027$). | 283 |
| B.9 | Best model found for H1*-A3* DID score (Conditional $R^2 = 0.098$). | 284 |
| B.10 | Best model found for H2*-H4* DID score (Conditional $R^2 = 0.013$). | 284 |
| B.11 | Best model found for CPP DID score (Conditional $R^2 = 0.073$). | 285 |
| B.12 | Best model found for CQ DID score (Conditional $R^2 = 0.077$). | 285 |
| B.13 | Best model found for PIC DID score (Conditional $R^2 = 0.13$). | 286 |
| C.1 | Male American English models' metadata. | 290 |
| C.2 | Female American English models' metadata. | 291 |
| C.3 | Male French models' metadata. | 292 |
| C.4 | Female French models' metadata. | 293 |

| | | |
|-----|---|-----|
| C.5 | Number of hours/week spent reading, writing, speaking, and listening to English reported by each subject. Subjects also specified whether they would listen more often to British or American English varieties. | 294 |
| C.6 | Each subject reported the amount of time they spent in an English-speaking country in months (1 week = 0.25 months). They self-reported their speaking proficiency on a scale from 1 (poor) to 7 (high), as well as their degree of accentedness on a scale from 1 (weak) to 7 (strong). The column "Accent" refers to whether they thought they had more of an American or British accent. | 295 |

PERCEPTION AND ACCOMMODATION
AMONG FRENCH LEARNERS OF ENGLISH:

An acoustic and electroglottographic study of creaky voice

Introduction

The human voice is a fascinating tool evolving from a very complex sequence of cognitive, physiological, and acoustic events, which is often referred to as the *speech chain* (Denes & Pinson, 1993). It results from the coordinated actions of the respiratory system, laryngeal (e.g. vocal folds), and supralaryngeal articulators (e.g. tongue, lips)¹. These actions generate an acoustic signal which travels to reach the eardrum, causing it to vibrate so that we perceive sounds. The voice carries long-term physical (e.g. sex, age), social (e.g. ethnicity, geographical background), and psychological (e.g. psychological states, emotions) characteristics. More precisely, “[...] there is something about the individual’s voice that is indicative of his personality” (Sapir, 1927: 896). The 'auditory face' of speakers is unique, like a fingerprint, which allows the formation of person-specific representations for a particular voice (Lee *et al.*, 2019). It is very surprising and rare to associate the voice of a person you hear talking for the first time to that of a person you already know. One of the linguistic elements responsible for listeners to recognize a person by the sound of their voice is voice quality, which is similar to phonation type in voice quality theory (Esling, 2019). Phonation types are laryngeal configurations that lead to the auditory-perception of different voice qualities (production vs. perception). Known voices are easily recognized and listeners can form a rapid and distinct impression of a person they do not know simply from their voice. As reported by Kreiman *et al.* (2008), the human ability to form an impression of a speaker’s voice arises from a long evolutionary process: many different animal species use vocal qualities to signal threat, size, or relationships. Impressions one make from a voice can, however, be very inaccurate: the mental picture formed based from hearing a voice can clash. Everyone has once thought “Huh, I wouldn’t have imagined them looking like that” when seeing someone

¹It has also been narrowly defined as the sound that is produced by the vibration of the vocal folds (Kreiman *et al.*, 2008). In this view, the voice only corresponds to the linguistic [+ voicing] feature and excludes all other parameters occurring during speech production (e.g. the effects of vocal tract resonances, turbulence noise, etc.).

for the first time but after hearing them talk on the phone, for instance. Impressions formed upon hearing a voice can be extremely problematic for persons suffering from voice disorders. Kreiman *et al.* (2008) reported that patients with a voice disorder suffer from the fact that their disordered voice does not convey their real self. In extreme cases, those patients end up avoiding speaking, leading to social and work-related difficulties. Pathologies, but also small changes in social context, psychological state, or emotion, can cause significant variability in an individual's voice. These changes can be easily perceived by listeners. In French, for instance, it is common to say to a person that they have *une petite voix* when they sound a bit off. Voice, and voice quality, can also be purposely modulated; we can play with it depending on our intentions (e.g. to signal irony or sarcasm). There are many different voice qualities existing in the world's languages, and their function may vary in different linguistic ways. In this thesis we will focus on one specific phonation type/voice quality called *creaky voice*.

Creaky voice, sometimes referred to in the literature as *vocal fry*, *creak*, *glottalization*, or *laryngealization*, is commonly said to be produced when the vocal folds vibrate very slowly and spend more time approximated than apart (Gobl & Ní Chasaide, 1992, 2010; Johnson, 2011; Podesva & Callier, 2015), resulting in low f_0 and low airflow rates (Podesva, 2013)². Listeners can hear the separate vocal fold vibrations, giving the perceptual impression of “running a stick along a fence, or slowly opening a door with creaky hinges” (Biemans, 2000: 27)³. The sexual dimorphism in the vocal anatomy of humans (e.g. length of the vocal tract) implies that men speak with a lower voice than women. Creaky voice was, therefore, primarily associated with male speech. Male speaking with a creaky voice were perceived as being more authoritarian, intelligent, and self-confident (Esling, 1978; Yuasa, 2010). The idea that low-pitched voices were better perceived than high-pitched voices actually dates back to the 20th mid-century. As mentioned by Austin (1965: 37), “low pitch has lately become fashionable for women, but fifty years ago all 'ladies' spoke with a high pitch”. Yuasa (2010) reported that middle-aged women sometimes ask laryngologists how they could lower their voices because their normal speech sounded awfully 'screechy' to them. Women might have intuitively realized that their socio-

²It was formerly classified as being pathological, as a clinical syndrome associated with abnormal laryngeal output (Hollien *et al.*, 1966).

³We will see that there actually exist sub-types of creaky voice, each having their own set of characteristics.

professional success depended on assimilation within the dominant group, hence within men (Yuasa, 2010). The adaptation of low pitch undoubtedly led to the adaptation of creaky voice. Several studies showed that creaky voice usage extensively increased in young female American speech these past few decades (Lefkowitz & Sicoli, 2007; Yuasa, 2010; Podesva, 2013). Yuasa's (2010) monograph precisely focused on this phenomenon and its spread. She observed that young Northern Californian women speaking with a creaky voice were perceived as being urban, well-educated, and to be highly positioned in the social hierarchy (Yuasa, 2010).

Both studies I conducted during my master's degree (Burin & Ballier, 2017; Burin, 2018) focused on phonetic accommodation between French and native speakers of English. As stated by Dufour & Nguyen (2013:1), "Imitation is an all-pervading process by which individuals adjust to one another in social interaction, and is seen as one of the fundamental mechanisms of human development". Humans are said to be hardwired to imitate (Coles-Harris, 2017), and imitation plays a particularly important role in language acquisition. It is the first process in which children engage to develop their ability to speak, but it is also very important in second language acquisition. When analysing one corpus I realised that female French advanced learners of English who would have a more pronounced American accent would use quite a few instances of creaky voice. Many questions started popping up in my mind: is this the result of adaptation due to exposure to the English language or is it related to idiosyncrasies present in the L1? Does creaky voice exist in French? How is creaky voice perceived by French speakers? Has creaky voice more prestige when produced by French or American English speakers? Considering that gender asymmetry has been observed in many convergence studies, would speakers converge more towards male or female creaky voice? We eventually came up with the following research question:

How are the evaluation and accommodation of creaky voice shaped by language-specific characteristics, and can they be influenced by social evaluation?

Many (cross-linguistic) studies have been conducted on linguistic and phonetic accommodation, but convergence in voice quality has received much less attention. To our knowledge, this study is the first on accommodation of creaky voice from French speakers towards both French and American English speakers. It lies at the intersection of many disciplines: physiology, acous-

tics, psycholinguistics, and sociolinguistics.

A two-fold approach was adopted, which includes both acoustic and electroglottographic measures of convergence and a perceptual evaluation of voice quality. We provided analyses on multiple acoustic and electroglottographic measures to observe whether female French learners of English would align on the same dimensions to both native speakers of French and American English. There has been many studies including listener judgments on American English creaky voice (e.g. Yuasa, 2010; Anderson *et al.*, 2014, Lee, 2016), but little is known as to how it is perceived by non-native speakers of English. We tested the influence of language and gender on creaky voice accommodation and perception. We expected more convergence towards American English than towards French creaky voice, and more towards male than towards female model speakers. More positive evaluation towards American English creaky voice than towards French creaky voice was also expected.

Speech stimuli were elicited from 12 native speakers of both American English and French (3 male, 3 female, in each language). Audio and electroglottographic (EGG) signals were recorded simultaneously. Electroglottography is a non-invasive technique that allows the observation of the properties of the vibrating vocal folds during phonation. Ten acoustic and two EGG measures were extracted and analysed. Recordings took place at the University of Washington (Seattle, WA)⁴ and Université Paris Cité. Ten short (5-7 syllables) sentences were produced with either word-final creak or without word-final creak by each model speaker, in their native language. Only declarative sentences were used for they elicit a low tone in both languages. We focused on the last accented word of the prosodic unit to have a balanced sample between the two languages. The last accented word is a monosyllabic word containing either a low or a high vowel⁵. Creaky voice is not so a common feature in French than it is in American English and, to the best of our knowledge, little is known as to how its acoustics may differ across language. Phonatory settings in French are described as “nasalized, breathy and sometimes whispery but apart from hesitation manifestations, there is no mention of a possible creaky voice phenomenon in this language” (Benoist-Lucy & Pillot-Loiseau, 2013: 2395). We first

⁴Research visit to UW funded by the Graduate School and Labex EFL in 2019.

⁵Stimuli were controlled to retain as much balance as we could across the two languages. Unfortunately, due to the number of constraints we had, we were unable to control for word frequency although previous findings showed that lexical frequency significantly influenced convergence effects (Pardo *et al.* 2013, 2017).

conducted an acoustic and electroglottographic analysis that provided an objective comparison between French and American English creaky voice. Convergence was then studied from a multi-dimensional perspective. 20 cisgender female native speakers of French aged 20-30, all majoring in English at Université Paris Cité were recruited to take part in the experiment. The experiment consisted in three different tasks: a reading task, a repetition task, and a rating task. Subjects first conducted the reading task in which they read the same stimuli as those recorded by the model speakers. This served as their baseline production. Each subject was then presented with the different auditory stimuli in each language. They had been instructed to repeat and imitate as closely as they could the production of the model speakers beforehand. Although they were explicitly instructed to imitate what they heard, we will use the terms imitation and convergence interchangeably along this study for we do not know what speech features (e.g. intonation, accent, speech rate, etc.) they actually perceived and managed to imitate, and whether creaky voice was one of them. Accommodation patterns were observed by comparing data from the reading and the repetition/imitation task. Between-language and gender-based differences within-language were analysed. A judgment rating task was then conducted in which female French learners of English were asked to evaluate their impression of the model speakers' voice on a 4 six-point semantic scale (i.e. pleasant, attractive, powerful, educated). They listened to 3 same sentences produced with either word-final creak or without word-final creak produced by 8 (2 male, 2 female, in each language) model speakers. Linear mixed-effects models were built to compare French and American English creaky voice, as well as to assess convergence on acoustic and EGG dimensions. They were carried out individually for each dimension. Ordinal logistic regressions were performed to compare evaluations of creaky vs. non-creaky voice.

This thesis is organised as follows. In Chapter I we provide an overview of the process of phonation. We describe the different muscular and cartilaginous structures that form the larynx and how laryngeal configurations can be varied to produce different phonation types in section I.1. Acoustic measurements that are commonly used in phonation studies are reported in section I.2. The different functions of non-modal phonation types are reviewed in section I.3.

In Chapter II we discuss the process of phonetic accommodation. The different linguistic levels at which convergence has been observed and popular experimental designs used to assess

convergence are reviewed in section II.1. From a psycholinguistic perspective, convergence is said to result from an unmediated link between perception and production. We report theories supporting this position and include a section on cross-linguistic accommodation in section II.2. Models of L2 perception and acquisition, as well as effects of proficiency and 'phonetic talent' are presented. From a sociolinguistic perspective, convergence is said to be driven by external and social influences. We review the effects of gender, interactional dominance, cultural affiliation, and social preference on convergence in section II.3. A summary of the very few studies on creaky voice alignment is provided in section II.4.

In Chapter III we describe the whole experimental design of our study. We detail how recordings were made and what electroglottography is in section III.1. We describe the speech stimuli used in this study and how they were elicited in section III.2. Section III.3 deals with participant selection. We discuss the language test we used to assess the participants' proficiency level and provide a detailed description of the three tasks we conducted (a reading, a repetition/imitation task, a judgment rating task) in section III.4. Our hypotheses and the different measurements (acoustic and EGG) we made are reported in section III.5 and III.6, respectively.

Analyses and results are divided into two chapters: In Chapter IV we provide a descriptive analysis of creaky voice across language and gender. We discuss data processing and visualisation, as well as statistical model selection, in section IV.1, IV.2. and IV.3. We analysed each acoustic and EGG measures we made on the data collected from the recordings of speech stimuli in section IV.4 and IV.5. The aim was to compare the acoustics of American English and French creaky voice, and to observe any gender-based differences within-language. We conducted principal component analyses to better understand what variables were responsible for most variance in the production of creaky voice in our dataset, and whether we observed any variation in the way variables were correlated across language. Results are reported in section IV.6.

In Chapter V we report the results of the different tasks we conducted. A comparison of creaky voice usage across language in the reading task is provided in section V.1. Convergence effects observed in the repetition/imitation are analysed in section V.2. We provide a description of how convergence was measured and an analysis of convergence on every acoustic and EGG measures we conducted. Ratings of creaky VS non-creaky voice are reported in section V.3. All

convergence and rating analyses include the effect of language and gender effect within-language.

We conclude this thesis in Chapter VI. We review our hypotheses and qualify our results, but also our lack of results. We discuss how this study could be improved and further developed. We reevaluate our protocol and outline further research.

CHAPTER I: Phonation

Phonation, in its narrow definition, is the process by which the air expelled from the lungs sets the vocal folds into vibration. Its periodic variations result in a quasi-periodic waveform which reflects the rapid opening and closing of the vibrating vocal folds. Supraglottal articulators (i.e. the tongue, lips, and nasal cavity) and their multiple configurations modulate that sound wave to produce different speech sounds (Gobl & Ní Chasaide, 2010). Speakers can also control different muscles and cartilages that alter the shape (length and thickness) of the vocal folds, as well as their separation. These different laryngeal configurations result in the production of different phonation types/voice qualities. Humans have developed the ability to detect these very subtle changes in voice quality, and interpret their meaning and function that may vary across language (Gobl & Ni Chasaide, 2010).

CHAPTER CONTENT

This chapter provides an overview on the process of phonation. In the first section we discuss the physiology of voice production. The muscular and cartilaginous structures that form the larynx are presented. It will help understand how different laryngeal parameters, when combined in different ways, lead to the production of various phonation types. The laryngeal settings involved in these phonation types are then described. In the second section we review the acoustic measurements that are commonly used to assess phonation types. The different functions of non-modal phonation types are reported in the third section. We will see that non-modal phonation types can be both linguistically and socially meaningful.

1 Physiology and phonation types

We provide an overview of the anatomy and physiology of the larynx in section 1.1. The physiological correlates of the different phonation types are discussed in section 1.2. In section 1.3 we describe the most common phonation types.

1.1 Anatomy and physiology of the larynx

1.1.1 General description

Phonation occurs in the larynx, a complex structure made of numerous cartilages and muscles. It houses the vocal folds, that are two muscles modulating the flow of air being expelled during phonation. The space between the vocal folds is often referred to as the *glottis*. According to the myoelastic-aerodynamic theory of phonation developed by Van den Berg in 1958, the closed glottis resists the flow of air coming from the lungs. As a consequence, the subglottic pressure increases until overcoming this resistance, leading to the opening of the glottis. As soon as the glottis is open, the subglottal pressure decreases, and the vocal folds snap shut again. It only takes one second for the vocal folds to adjoin and split a hundred of times (Henrich, 2001). This vibratory movement generates acoustic waves that spread inside the vocal tract that acts as a filter.

1.1.2 Cartilaginous structure

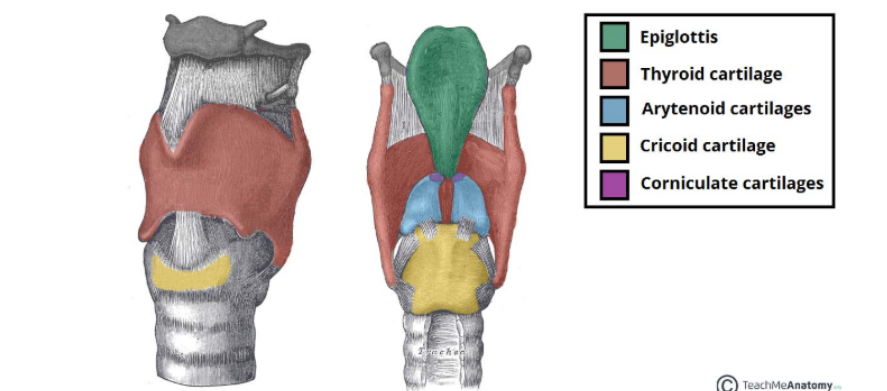


Figure 1: Cartilaginous structure of the larynx (from: Jones & Barnes, 2019).

The different cartilages involved in phonation are reviewed after Jones & Barnes (2019)⁶. Figure 1 illustrates the different cartilages that are present in the larynx. The thyroid, the cricoid, and the two arytenoids are particularly important in the process of phonation. For this reason, only these cartilages are detailed.

The thyroid cartilage

The thyroid cartilage is the largest structure in the larynx. It is composed of two sheets that meet on the anterior side to form the laryngeal prominence often called the *Adam's apple*.

The cricoid cartilage

The cricoid is the only complete circle of cartilage present in the larynx. It forms its base and is said to resemble a signet ring in shape. It provides an attachment for the inferior horns of the thyroid cartilage and articulates with the arytenoids posteriorly.

The arytenoid cartilages

The arytenoids are pyramidal shaped structures that sit directly on the cricoid cartilage. They connect with several other structures. Their apex articulates with the corniculate cartilage, and their base with the superior border of the cricoid cartilage. The arytenoids are made of two processes. The vocal process provides attachment for the vocal ligament, which is part of the vocal folds. The muscular process provides attachment for the lateral and posterior cricoarytenoid muscles, described below. When breathing, the arytenoids are spread apart, allowing the airflow to circulate without any constraint. At the beginning of phonation, those same cartilages spin to adjoin and allow the glottis to close (Henrich, 2001).

1.1.3 Muscular structure

In addition to the cartilages we have just seen, the larynx is made of different muscles: the extrinsic muscles, and the intrinsic muscles (Henrich, 2001).

- The extrinsic muscles (*sub- and sus-hyoid muscles*⁷): they act to move the larynx superiorly

⁶Available as: <https://teachmeanatomy.info/>

⁷Muscles attached to the hyoid bone which is a 'U' shaped structure located at the base of the mandible (Jones & Barnes, 2019).

and inferiorly, as well as acting on the movement of the lower jaw – also called the mandible – by lowering or raising it.

- The intrinsic muscles (*thyroarytenoid, interarytenoid, cricothyroid, and cricoarytenoid muscles*): they act directly on the vocal folds and, more specifically, on their position, length, and tension, as well as on their vibratory movement.

Only the intrinsic muscles will be detailed here for they are specifically important in phonation. They are illustrated in Figure 2. The extrinsic muscles are primarily involved in other functions such as swallowing.

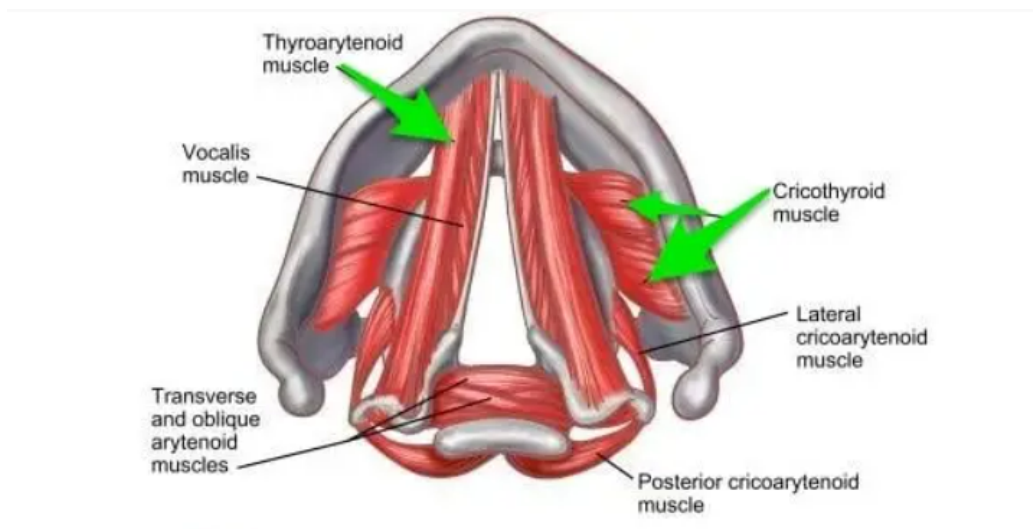


Figure 2: Intrinsic muscles of the larynx (from: The Anatomical Chart Series, 1993).

The thyroarytenoid muscles

The thyroarytenoid muscles (TA) connect the thyroid cartilage to the vocal processes of the arytenoid cartilages. The lateral part of the thyroarytenoid muscle is often referred to as the *vocalis*. It is the main portion of the vocal folds. The thyroarytenoid muscles, when contracted and unopposed, relax and shorten the vocal folds. They also help the closing of the glottis by drawing the arytenoids towards the thyroid cartilage (Gick *et al.*, 2013).

The interarytenoid muscle

The interarytenoid muscle (IA) connects the two arytenoid cartilages. It is composed of the oblique interarytenoid muscle, and of the transverse interarytenoid muscle. The oblique interarytenoid muscle goes diagonally from the top of one cartilage to the bottom of the other. The transverse interarytenoid muscle goes horizontally between the posterior surfaces of the two arytenoid cartilages. This muscle serves in the approximation of the arytenoid cartilages (Gick *et al.*, 2013).

The cricothyroid muscle

The cricothyroid muscle (CT) links the cricoid and thyroid cartilages. Its contraction increases the distance between the two cartilages, contributing to increase the tension on the vocal folds (Gick *et al.*, 2013).

The cricoarytenoid muscles

There are two different cricoarytenoid muscles: the lateral cricoarytenoid muscles (LCA), and the posterior cricoarytenoid muscles (PCA). The main function of the LCA muscles is to rotate the arytenoid cartilages towards the interior, resulting in the approximation of the vocal folds. The reverse process, therefore the separation of the vocal folds, will be possible when the arytenoid cartilages rotate towards the exterior. It is achieved under the action of the PCA muscles.

1.1.4 Differences between men and women

From the onset of puberty until roughly 20 years of age, the female and male larynx undergo different growth patterns. The larynx is lower in the neck of the adult male, for whom the average length of the vocal tract (region comprised between the vocal folds and lips) is approximately 17-18 cm, as compared to that of the adult female which is approximately 14.5 cm (Simpson, 2009). The length of the vocal folds also differ. It is comprised between 13 to 17 mm for women and between 17 to 24 mm for men (Titze, 1994; Childers, 2000). The medial surface of the vocal folds appears to be curved in men, as a result of tissue bulging, whereas women's vocal folds are more triangular shaped.

1.2 Physiological correlates of phonation types

There are numerous ways in which the state of the larynx can be manipulated to give rise to sound sources with distinctive physiological properties. According to Laver (1980), there exist three laryngeal parameters that, when combined in different ways, lead to the production of different phonation types. These three parameters are known as *longitudinal tension*, *adductive tension*, and *medial compression*. All these parameters, illustrated in Figure 3, are determined by actions of the muscles and cartilages previously described.

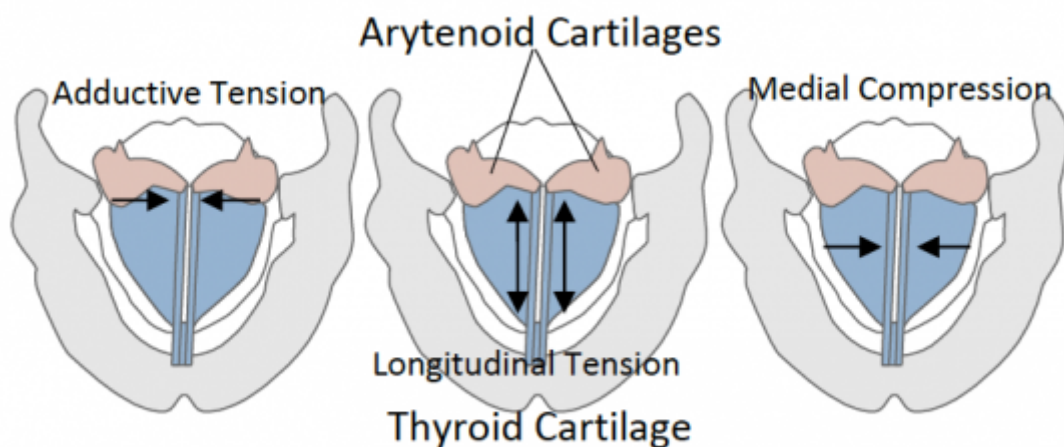


Figure 3: Laryngeal parameters (from: Wright *et al.*, 2019).

Longitudinal tension

Longitudinal tension represents the tension of the vocal folds. It is primarily controlled by the TA muscle. Its contraction curtails longitudinal tension by shortening the length of the vocal folds when unopposed. This causes the vocal folds to have more mass per unit length, thus to vibrate more slowly, resulting in a lower fundamental frequency (Raphael *et al.*, 2007). Conversely, when opposed, vocal fold tension increases (Zemlin, 1998). The CT muscle can also have an impact on the tension on the vocal folds, therefore on their length. Its contraction increases the distance between the two cartilages it connects, resulting in an increase in longitudinal tension (Zemlin, 1998).

Adductive tension

Adductive tension has been defined as the force by which the arytenoids are drawn together. It is controlled by the IA muscle that, when contracted, induce the approximation of the vocal folds (Gobl & Ní Chasaide, 2010).

Medial compression

Medial compression the force by which the vocal folds are brought together, through the approximation of the vocal processes of the arytenoids (Gobl & Ní Chasaide, 2010). It is primarily controlled by the LCA muscles so that, when rotating the arytenoid cartilages inward, will bring the vocal folds together (Zemlin, 1998).

1.3 Phonation types

The three laryngeal parameters detailed in the previous section act together and can be combined in different ways to produce various phonation types⁸. Phonation types across language have been characterised in terms of a continuum which results from the variations of the degree of aperture between the arytenoid cartilages. Spread and constricted glottis refer to the 'endpoints' (Ladefoged, 1971; Gordon & Ladefoged, 2001). This phonation continuum is illustrated in Figure 4.

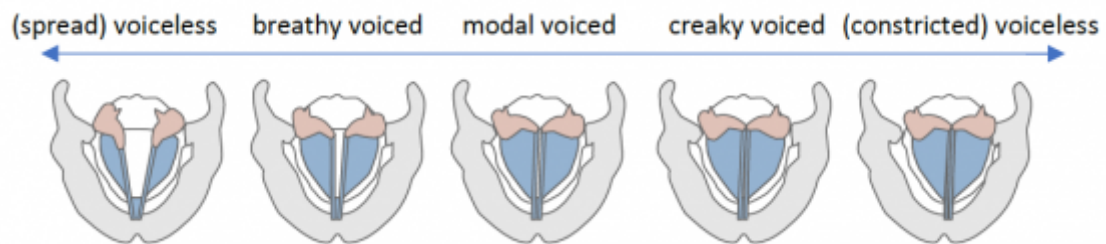


Figure 4: Phonation types and their relationship to glottal aperture (from: Wright *et al.*, 2019).

This continuum can be considered as being oversimplified for the list of all existing phonation types appears to be more exhaustive, which means that much more complicated laryngeal

⁸Our study being primarily concerned with the state of the glottis, we here consider the narrow definition of phonation types. However, laryngeal configurations also involve the muscular structures of the tongue and pharynx, for instance (e.g. Edmonson & Esling, 2006; Esling *et al.*, 2019).

configurations need to be taken into account to interpret voice quality variations. In this section we only review the phonation types included in the continuum⁹. It also has to be borne in mind that researchers have not yet reached a consensus regarding the labels used to define the different phonation types:

[...] a given label may refer to different phenomena while different labels may be used to describe very similar phenomena, depending simply on the user's understanding of the term.

Gobl & Ní Chasaide (2003: 192)

An overview of the laryngeal settings involved in each phonation type included in the continuum is given below. An emphasis will be put on creaky voice for this it is the main topic of this study .

[spread] Voiceless

This phonation type is realised with neither medial compression nor adductive tension. As a result, the vocal folds are completely separated and cannot be set into vibration as air passes between them (Raphael *et al.*, 2007; Johnson, 2011).

Breathy voice

Breathy voice is produced with little longitudinal tension, medial compression, and adductive tension. The glottis is kept open along most of its length, vibrating without ever fully closing (Gobl & Ní Chasaide, 2010; Podesva, 2013). More specifically, “the arytenoid cartilages are well separated at the back but the vocal processes are sufficiently approximated so that the vocal folds vibrate when a lung pressure is applied to the system” (Klatt & Klatt, 1990: 822). Airflow is highly important due to the large average glottal opening (Garrelek, 2014), resulting in considerable audible friction noise (Gobl, 1989; Gobl & Ní Chasaide, 1992, 2010; Podesva & Callier, 2015).

⁹The reader can, however, refer to Laver (1980) and Esling *et al.* (2019) for a general description of the main phonation types; Keating (2014) for a detailed description of falsetto voice; Gobl & Ní Chasaide (1992) for a detailed description of whispery, lax, and tense voices; or Wendhal (1963) for a detailed description of harsh voice.

Modal voice

Modal voice (unmarked voicing) is considered as the 'default' from which all other phonation types vary (Epstein, 2002). Its configuration is achieved with moderate adductive tension and medial compression, and little longitudinal tension (Laver, 1980; Gobl, 1989). It is characterised by an average glottal opening (Garellek, 2014) that allows complete closure during glottal periods (Garellek & Keating, 2011), resulting in no audible aspiration noise (Gobl & Ní Chasaide, 2010). The vocal folds open widely and are tightly adducted along the anterior margins (Blomgren *et al.*, 1998; Johnson, 2011). This produces maximum vibration, which is, itself, due to the fact that the ratio of time the vocal folds spent approximated and apart is approximately proportional (Klatt & Klatt, 1990).

Creaky voice

Creaky voice, which is frequently referred to in the literature as *vocal fry*, *creak*, *glottalization* or *laryngealization*, involves low subglottal pressure, strong adductive tension and medial compression, but very little longitudinal tension (Gobl, 1989; Gobl & Ní Chasaide, 1992; Drugman *et al.*, 2013). It is commonly said to be produced when the vocal folds are approximated and shortened (Yuasa, 2010) as the arytenoid cartilages are drawn together. Because of the high adductive tension, vibration only occurs on the anterior side, hence away from the arytenoids (Ladefoged & Johnson, 2015). The vocal folds margins remain flaccid because of decreased longitudinal tension (Zemlin, 1998). The vocal folds are very slowly vibrating and spend more time approximated than apart (Gobl & Ní Chasaide, 1992, 2010; Johnson, 2011; Podesva & Callier, 2015). As a result, there is a very short open period followed by a very long period during which the vocal folds are completely approximated (Blomgren *et al.*, 1998). Mean airflow rate is very low (Catford, 1964; Podesva, 2013).

More recently, researchers found that these characteristics were not observed in all instances of creaky voice, and that specific sub-categories of creaky voice existed, each with its own set of characteristics. These different types of creaky voice are reported below, based on Keating *et al.* (2015) and Wright *et al.* (2019). They are known as prototypical creaky voice, vocal fry, multiply pulsed voice, aperiodic voice, non-constricted creak, and tense/pressed voice.

1. Prototypical creaky voice corresponds to the creaky voice described above. It is char-

- acterised by increased adductive tension and medial compression, but low longitudinal tension, which results in an irregular signal (irregular f_0) with a lower vibratory rate. The glottis is constricted and the vocal folds are close together, with a small peak glottal opening and long close phase.
2. Vocal fry is characterised by high adductive tension, moderate medial compression, and low longitudinal tension. The glottis is constricted and individual pulses are distinct and separately audible. The vibratory rate (hence f_0) is low, but not necessarily irregular (even often quite periodic).
 3. Multiply pulsed voice is characterised by high adductive tension, longitudinal tension, and medial compression¹⁰. Alternation of high/low (alternating amplitude) and long/short pulses (alternating duration) are observed within one cycle (Gobl, 1989; Ladefoged & Maddieson, 1996), which is due to the presence of ventricular incursion. The ventricular folds push down and cover the 'true' vocal folds, which causes an increased mass. As a consequence, the frequency of vibration is lowered, and secondary vibrations may occur (Edmondson & Esling, 2006). The glottis is constricted and the presence of noise can be perceived (percept of roughness). In the case of double pulsing, there are two simultaneous periodicities, leading to multiple f_0 (these pulses generally have a long closed phase), or indeterminate f_0 .
 4. Aperiodic voice is produced with increased adductive tension and low longitudinal tension. Extreme periodic irregularity is present (beyond irregular = no periodicity), which results in perceived noise and no perceived pitch.
 5. Non-constricted creak (or Slifka voice)¹¹ has an irregular vibratory rate, and is produced with low longitudinal and adductive tension, but with high medial compression. This results in a signal with a perceived irregular and low pitch, as well as turbulence perceived as breathiness. The glottis is spread, not constricted. Therefore, airflow circulating through the glottis is higher, not lower.

¹⁰Personal communication with Patricia Keating.

¹¹Based on the work of Slifka (2000, 2006).

6. Tense/pressed voice is produced with a constricted glottis and involves a high degree of tension in the entire vocal tract. Fundamental frequency is, however, neither irregular nor low (can be mid or high and regular).

As mentioned earlier, there has been some overlap or differences in the terminology used by researchers regarding creaky voice. The term 'creaky voice' may then sound ambiguous for it represents a cluster of phonation types/voice qualities. We will use that term in our study as the representation of the abstract phonological category, in the sense that it may be realised differently by different speakers, in different languages, and/or in different linguistic environments.

[constricted] Voiceless

This phonation type involves high adductive tension and medial compression. The glottis is closed by tight contraction of the interarytenoid and cricoarytenoid muscles. This closure prevents air from passing through the glottis, resulting in the production of a voiceless stop, also called *glottal stop* (Gick *et al.*, 2013).

Although the phonation types mentioned above are characterised by different laryngeal settings, it should be borne in mind that they vary in a continuous and not a categorical way. Any voice quality may occur at different rates across speakers of one language/dialect, and across language. There is no 'absolute' breathy or creaky voices, but some voice qualities can appear as being breathier or creakier than others. This is partly due to their relative differences that many names for voice qualities exist, as we have seen in the case of creaky voice. As mentioned by Esling *et al.* (2019: 15), “varying degrees of whisperiness, harshness, and creakiness can be combined, all being functions of laryngeal constriction. The descriptive convention is usually to specify whether the constituent elements are present to a slight, moderate, or extreme degree. Thus, a voice could demonstrate varying degrees of the constituents of *harsh whispery creaky voice*.”

In this section we discussed the physiology of the larynx and how different laryngeal configurations give rise to the production of various phonation types. The next section illustrates the fact that there is a direct cause-and-effect relationship between physiology and acoustics. The

most popular acoustic measures used by researchers in this field are reviewed in the next section. Other measures such as jitter, shimmer, or subharmonic-to-harmonic ratio (see Panfili, 2018), can also be used, but will not be detailed here.

2 Acoustic measures of phonation types

As seen previously, one model that has become very popular to describe phonation type contrasts is the continuum of glottal stricture (Ladefoged, 1971; Gordon & Ladefoged, 2001). One of the reasons for the popularity of this model is that there exists a relationship between measures related to the average glottal opening and acoustics (Garellek & Keating, 2011). Yet this uni-dimensional model appears to be insufficient: other laryngeal configurations have to be considered than just the degree of glottal stricture. The literature has, however, not yet agreed on one single set of acoustic properties that could be used to differentiate between all phonation types. It is even possible that an agreement will never be reached. Keating *et al.* (2010), for instance, found that different sets of acoustic properties better distinguished between phonation types in four different languages that use phonation phonemically (Jalapa Mazatec, White Hmong, Southern Yi, and Gujarati). This observation suggests that the complexity of phonation is even greater, for different languages appear to produce phonation types in different physiological ways. It is also very likely that all speakers of a same language do not produce phonation types in the same way, especially in languages in which phonation is not phonologically contrastive. Nonetheless, many studies suggest that a particular set of acoustic properties can be useful in measuring and comparing phonation types. The following is an overview of this set of acoustic measures.

Fundamental frequency (f_0)

Fundamental frequency (f_0) is controlled by the rate of vibration of the vocal folds and is the perceptual correlate of pitch. f_0 is defined as $1/T_0$, where T_0 is the fundamental period which, itself, corresponds to the duration of the glottal cycle as defined by the time between the main excitation of two consecutive pulses (Gobl & Ní Chasaide, 2010). A high f_0 is correlated with increased vocal fold length and tension, while a low f_0 is correlated with decreased vocal

fold length and tension (Laver, 1980). More specifically, pitch raise is mainly achieved through contractions of the CT muscle, which stretches the vocal folds, resulting in a decreased mass and increased stiffness. An increase in airflow also contributes to pitch rise. Conversely, a decrease in the activity of the CT muscle, and a decrease in airflow, lowers pitch (Hirose, 1997).

Intensity

The intensity of a sound is what listeners perceive as loudness. It represents the amount of energy in the acoustic signal and is correlated with the amplitude of the wave. It is mostly controlled by the subglottal pressure (Zhang, 2016) and is especially important for stress (Gordon & Applebaum, 2010).

Spectral tilt

Spectral tilt reflects the “the degree to which intensity drops off as frequency increases” (Gordon & Ladefoged, 2001: 399). It is an excellent indicator of the degree of gradualness/abruptness of vocal fold closure (Avelino, 2010), and is considered as the most reliable correlate of increased constriction or spreading during voicing. A decreased spectral tilt indicates constriction, while an increased spectral tilt indicates spreading (Klatt & Klatt, 1990; Gordon & Ladefoged, 2001; Hanson *et al.*, 2001; Kreiman *et al.*, 2012; Keating *et al.*, 2015). The main spectral measures are:

- $H1^*-H2^{*12}$ is the most widely used spectral measurement. It corresponds to the amplitude difference between the first harmonic (the fundamental) and the second harmonic in the spectrum. It is a correlate of the open quotient (OQ), which represents the ratio of the duration of the open phase to the duration of a complete glottal cycle (Hanson & Chuang, 1999). OQ is a common measure in studies using electroglottography. It has been considered by Henrich *et al.*, (2005) as a dimensionless parameter, ranging from 0 (no opening) to 1 (no or incomplete closure). Essentially, the smaller the amplitude of the second har-

¹²Harmonic amplitudes are affected by the vocal tract filter and the source function. Asterisks indicate that measurements were corrected for local formant frequencies and bandwidth influences to allow comparisons of those measures made across different speakers and different vowels. As mentioned by Hanson & Chuang (1999: 1066), “it is very important to make this correction when comparing the acoustic measures across different vowels, for which F1 can considerably varies” For H1-H2, for instance, only F1 and F2 are used in the correction, but for H1-A3, F3 also has to be used.

monic relative to that of the fundamental, the less abrupt the glottal closing gesture, and the greater to 1 is the OQ (Stevens, 1977; Holmberg *et al*, 1995).

- $H1^*-A1^*$ represents the amplitude of the first harmonic relative to the amplitude of the first formant (F1). It is said to reflect the bandwidth¹³ of F1, which indicates the presence of a posterior glottal opening (Hanson & Chuang, 1999). The presence of a posterior glottal opening (glottal chink, incomplete closure) leads to an increase of formant bandwidths (particularly F1) due to additional energy loss at the glottis, to an increase in the spectral tilt at higher frequencies, and to the generation of turbulence noise in the vicinity of the glottis (Hanson & Chuang, 1999).
- $H1^*-A2^*$ represents the amplitude of the first harmonic relative to the amplitude of the second formant (F2). It is related to the skewness of the glottal pulse (Avelino, 2010). The glottal pulse is usually skewed to the right, meaning that the opening phase tends to be longer than the closing phase. More specifically, pulse skewing affects the amplitude of low harmonics: the more symmetrical the pulse, the greater the amplitude of low harmonics (Gobl & Ní Chasaide, 2010).
- $H1^*-A3^*$ represents the amplitude of the first harmonic relative to the amplitude of the third formant (F3). It correlates with the ratio of the duration of the closed phase to the duration of a complete glottal cycle (Avelino, 2010). The mid- to high-frequency components are mainly influenced by how abruptly the airflow is stopped from circulating when the glottis closes (Hanson & Chuang, 1999).

Other spectral measures include $H1^*-H3^*$ (the amplitude of the first harmonic minus the amplitude of the third harmonic), $H2^*-H4^*$ (the amplitude of the second harmonic minus the amplitude of the fourth harmonic), $H4^*-2k^*$ (the amplitude of the fourth harmonic minus the amplitude of the harmonic closest to 2 kHz), or $2k^*-5k^*$ (the amplitude of the harmonic closest to 2 kHz minus the amplitude of the harmonic closest to 5 kHz).

¹³Formant bandwidths are related to the rate of acoustic energy loss in the vocal tract (Hanson, 1997). They reflect the degree of damping present, which is an indication of the degree of the glottal opening: “A high degree of damping is found where there is little or no closed phase in the glottal pulse as, for example, in breathy voice. Supraglottal factors also affect the degree of damping, and thus the formant bandwidths” (Gobl & Ní Chasaide, 2010: 395).

Harmonic-to-Noise Ratio (HNR)

Turbulence noise – sometimes referred to as aspiration noise – occurs in the vicinity of the glottis. It is a consequence of glottal opening. As stated by Hanson (1997: 474), “when the minimum glottal opening becomes larger [...], the spectrum amplitude of the periodic component becomes weaker at high frequencies [...], and the amplitude of the turbulence noise increases because of the increased flow”. Specifically, the more the vocal folds are spread, the more turbulent airflow is generated at the glottis. Low subglottal pressure can also increase the amount of noise, which is due to the fact that voicing is less regular and weaker in this condition (Garellek, 2019). The presence of aspiration noise can be measured using HNR (Kreiman *et al.*, 2014). This spectral measure represents the ratio of periodic to aperiodic components of the signal (Murphy *et al.*, 2008). HNR can be measured across different frequency bands (HNR05 = 0-500 Hz; HNR15 = 0-1500 Hz; HNR25 = 0-2500 Hz; HNR35 = 0-3500 Hz). A low HNR indicates more noise in the signal (Keating & Garellek, 2015). Consequently, HNR is lower in non-modal phonation, meaning that more noise than harmonics is observed.

Cepstral Peak Prominence (CPP)

CPP is a measure of the prominence of the peak of the cepstrum. It corresponds to the difference between the peak cepstral value and the mean of all cepstral values. A cepstrum is the result of the inverse Fourier transform of a log spectrum (de Krom, 1993). A more prominent cepstral peak indicates stronger and better-defined harmonic structures above the floor of the spectrum. This can be the result of a more periodic and less noisy signals, therefore, of little jitter or shimmer¹⁴ (Garellek & Keating, 2011). Essentially, it is a measure of periodicity: “a larger difference implies a greater ratio of periodic to aperiodic sound in the signal” (Blankenship, 2002: 180). CPP values are higher for modal than for non-modal segments (Garellek & Keating, 2011).

In this section we saw that a variety of acoustic measurements can be used to analyse phonation (and to differentiate between phonation types). Different acoustic measures have to be

¹⁴Measures such as jitter and shimmer have been used to quantify pulse-to-pulse variation. The term *jitter* also called 'frequency perturbation', refers to the aperiodicity of the fundamental frequency. The term *shimmer* or 'amplitude perturbation', refers to the aperiodic amplitude variation (Biemans, 2000: 28). Both *jitter* and *shimmer* can be analysed independently of CPP.

conducted simultaneously in phonation studies. One parameter analysed on its own may not enable perceptual discrimination between phonation types, for instance.

[T]he presence of a specific acoustic feature in the voice does not necessarily imply the perceptual impression of a certain voice quality. For instance, noise levels in the acoustic signal may be high without the speech giving the impression of breathy [...] quality.

Biemans (2000: 64)

In the next section we discuss the different functions of non-modal phonation types. We will see that they can be both linguistically and socially meaningful.

3 The different functions of non-modal phonation types

Previous sections have illustrated the fact that speakers have the ability to control the glottis to produce different speech sounds with a variety of phonation types. In this section, we will see that non-modal phonation types have different functions. They can serve a linguistic function in being phonologically contrastive, arise as allophonic variants, or be used to mark prosodic boundaries and prominence. They can also be varied to signal paralinguistic information on mood and attitude, and/or to serve a sociolinguistic function. Contrastive, allophonic, prosodic, paralinguistic, and sociolinguistic phonation will be discussed in section 3.1, 3.2, 3.3, 3.4, and 3.5, respectively.

3.1 Contrastive

Phonation types can be varied to produce phonological contrasts. Contrasts are mostly found among vowels (e.g. Jalapa Mazatec, Chong, Zapotec¹⁵), or sonorants (e.g. Kwakw'ala, Montana Salish¹⁶). Languages like Hindi, Tsonga, or Newar¹⁷ contrast breathy voiced and modal

¹⁵Jalapa Mazatec and Zapotec are Oto-Manguean languages spoken in the Mexican states of Oaxaca and Veracruz; Chong is an Austroasiatic language spoken in eastern Thailand.

¹⁶Kwakw'ala is Wakashan language spoken in western Canada.

¹⁷Tsonga is a Bantu language spoken in South Africa and Mozambique; Newar is a Sino-Tibetan language spoken in Nepal.

voiced consonants, more particularly among nasals, while languages such as Gujarati¹⁸ make this contrast mostly on their vowels¹⁹, as reported by Gordon & Ladefoged (2001). Creaky vowels contrast with modal vowels in numerous languages, and with both breathy and modal vowels in others. Jalapa Mazatec, for instance, is a dialect that possesses this three-way phonation contrast, in addition to a three-level tone contrast (low, mid and high), making it quite unusual (Garellek & Keating, 2011).

Phonation types can also be constrained to only occur with certain types of tone. This is the case in Southern Yi (Kuang 2011) in which the phonation contrast never occurs with a high tone, whereas the opposite has been observed in Northern Yi²⁰. In SADV Zapotec²¹ (Esposito, 2006), non-modal phonation types only occur with a falling tone. In Mandarin tones, creaky voice appears to be associated with the fourth falling tone and with the third-low dipping tone (Davison, 1991; Belotel-Grenié & Grenié, 2004).

3.2 Allophonic

In addition to being contrastive in some languages, non-modal phonation types commonly occur as allophonic variants of modal phonation. These variants appear to be mandated by linguistic structures and associated with particular environments. They are frequently observed in the vicinity of consonants that are not produced in modal phonation. More precisely, vowels adjacent to /h/ are often breathy in English (Klatt & Klatt, 1990), while vowels preceding voiceless stops are often creaky (Gordon & Ladefoged, 2001). This is due to the fact voiceless stops are often glottalized in coda position in English.

3.3 Prosodic

The presence of non-modal phonation in speech can also be associated with the prosodic location of segments. Both prosodic boundaries and prominence appear to have an effect on

¹⁸Indo-Aryan language spoken in western India.

¹⁹It should be mentioned that, even where vowels use contrastive creaky or breathy phonation, the breathiness or creakiness tends to be localized to a portion of the vowel only, and does not persist throughout. Non-modal phonation tends not to be extended over an entire segment but is rather confined to a portion of the sound and/or spills over onto an adjacent segment (Gordon & Ladefoged, 2001).

²⁰Sino-Tibetan languages spoken in southern and northern China, respectively.

²¹Oto-Manguean language spoken in Santa Ana del Valle in the Mexican region of Oaxaca.

voice quality variation.

3.3.1 Phrase boundary

The effect of phrase boundary on non-modal phonation has been considered in many studies. An increase in creak at phrase boundaries appears to be relatively common in American English, both phrase-initial and phrase-final (Umeda, 1978; Kreiman, 1982; Pierrehumbert & Talkin, 1992; Pierrehumbert, 1995; Dilley *et al.*, 1996; Redi & Shattuck-Hufnagel, 2001; Epstein, 2003; Ní Chasaide & Gobl, 2004; Slifka, 2006; Wolk *et al.*, 2012; Podesva, 2013; Garellek, 2014). Redi & Shattuck-Hufnagel (2001) observed a hierarchical effect for phrase-final creak in English, with creak occurring more frequently at higher prosodic boundaries. Lehiste (1965), Kreiman (1982), and Slifka (2006), observed creak as marking the ends of both paragraphs and sentences within paragraphs. The same pattern has been observed in some British English dialects. Henton & Bladon (1988) observed many occurrences of creak at the end of a single sentence produced in isolation. Laver (1980) suggested that RP speakers used creak as a signal of completion of their turn. This can also be explained by the fact that aerodynamic conditions are no longer favourable for modal voice to be used in this position. Utterance-final creak was also observed in other languages such as Czech (Lehiste, 1965), Serbo-Croatian (Lehiste, 1965), Swedish (Gobl & Ní Chasaide, 1988), and Finnish (Lehiste, 1965). Other non-modal phonation types can occur at the end of prosodic unit. Utterance-final breathy voice was observed in Finnish (Ogden, 2001) and Swedish (Gobl & Ní Chasaide, 1988). Klatt & Klatt (1990) found that declarative sentences may end in a breathy-laryngealized type of vibration in American English.

3.3.2 Prominence

In addition to phrase boundaries, prominence can also have an effect on phonation type. Klatt & Klatt (1990) found that manifestations of breathiness considerably increased for unstressed syllables in English. Creak seems to occur more frequently on stressed syllables in Swedish (Gobl & Ní Chasaide, 1988). Pierrehumbert & Talkin (1992) observed higher rates of glottalization on stressed-vowel initial syllables as compared to their reduced counterparts in English. Epstein (2002) found a strong effect of prominence on phonation type for prominent

words²² in American English. Prominent words were characterised by tenser phonation type than non-prominent words. She did not find a correlation between prominent syllables and phonation type, contradicting previous observations that creak appeared more systematically in accented than unaccented syllables (Pierrehumbert, 1995; Dilley *et al.*, 1996). She concluded that “voice quality is not a property of the stressed syllable, but a property of the word as a whole” (Epstein, 2002: 93).

3.3.3 Phrase boundary and prominence

Both prominence and phrase boundaries, when taken altogether, can also influence phonation type. Garellek (2014) observed that both prominence and phrase-initial position strongly favoured the presence of word-initial glottalization in American English and other languages. He argued that word-initial glottalization may arise from prosodic strengthening, a process by which sounds are more 'strongly' articulated in stronger prosodic positions (Garellek, 2014: 106). This corroborates Pierrehumbert & Talkin (1992), Pierrehumbert (1995), and Dilley *et al.*'s (1996) observations that word-initial vowels (even reduced vowels, but to a lesser extent) are more frequently glottalized at the beginning of intonational phrases, and when the word is pitch-accented. Epstein (2002) also observed that phrase-initial prominent words tends to have a tenser voice quality than phrase-final prominent words.

3.4 Paralinguistic

The notion of *paralanguage* was made popular by Trager G. L. in 1958. Researchers became interested in the idea that conversations consist of much more than just interchanges of spoken words, that there are other features accompanying speech.

Paralinguistic phenomena [e.g. gesture, facial expression, voice quality...] are non-linguistic elements in conversation. They occur alongside spoken language, interact with it, and produce together with it a total system of communication. They are not necessarily continuously simultaneous with spoken words. They may also be interspersed among them, or precede them, or follow them; but they are always integrated into a conversation considered as a complete linguistic interaction.

²²Prominent words are defined as the words bearing the most prominent pitch accent

Abercrombie (1968: 55)

Paralinguistic features, including voice quality, can be purposely varied to signal attitude, mood, or emotion (Crystal, 1997). Impressionistic observations have associated specific voice qualities with particular emotions or affective states. However, listeners tend to disagree on the mapping between voice quality and affect. One specific voice quality can be associated with numerous states, both with positive and negative valence.

[T]here is no one-to-one mapping between voice quality and affect: individual qualities appear rather to be associated with a constellation of affective states, sometimes related, sometimes less obviously related.

Gobl & Ní Chasaide (2003: 208)

Breathy voice was found to index authority in certain languages such as Zapotec (Sicoli, 2010), while it has been associated to intimacy in English (Laver, 1980). Grivičić & Nilep (2004) observed that, when the word “yeah” was pronounced with a creaky voice in American English, it would signal passive receptiveness indicating either disalignment between the interlocutors, or dispreference for the topic. Creaky voice in American English also appears to be used as a strategy to index surprise, admiration, or suffering (Ishi *et al.*, 2005), or as a mark of hesitation (Carlson *et al.*, 2006). Brown & Levinson (1987) proposed that creaky voice could be used to complain or commiserate among speakers of Tzeltal, a Mayan language spoken in Southern Mexico. Dillely *et al.* (1996) found that creaky voice is a frequent feature in the speech of female radio newscasters in American English. She suggested that it could be used to take an authoritative stance. This was supported by Lefkowitz & Sicoli (2007) who studied the speech of American college-aged women and found that they were more likely to use creaky voice when taking such authoritative stances. Mendoza-Denton (2007) studied the speech of Latina gang-affiliated girls in Northern California and found that they would use creaky voice most often when narrating fight stories. She hypothesised that creaky voice could index toughness. In American English, tense voice was associated with anger, joy, and fear in Scherer’s (1986) study, while it was associated with stress, anger, formality, and confidence in Ní Chasaide *et al.*’s (2004) study. Laukkanen *et al.* (1996) found that anger was characterised by a low OQ in Finnish, suggesting a rather tense

setting. Conversely, surprise, enthusiasm, and sadness tended to be characterised by a high OQ, suggesting a breathy setting. This corroborates Burkhardt & Sendlemeier's (2000) observations that tense voice tends to be associated with anger, and breathy voice with sadness in German. In RP English, a complete utterance with creaky voice has been observed as signalling 'bored resignation' (Laver, 1980).

The fact that listeners disagree on the mapping between voice quality and affect may be due to 1) cultural differences, 2) the type of judgment rating tasks, 3) and also to the fact that listener judgments of certain voice qualities are not solely based on the presence or absence of that voice quality, but rather on a combination of speech features that interact with one another in unique ways (Parker & Borrie, 2017). Moreover, Papcun *et al.* (1989) have argued that listeners who have had a life-long experience with voices develop central category constituents for vocal quality that they will, then, use to judge or remember voices they hear. These categories deriving from perceptual experience, listeners' background will affect their perceptual strategy when judging voices. This was confirmed by Kreiman *et al.* (1990) who observed that naive listeners would differ from experts in several dimensions to judge pathological and non-pathological voices.

3.5 Sociolinguistic

Voice quality can also have a sociolinguistic function in that it can be varied to differentiate between linguistic, regional, and social groups, as well as to manage personas.

3.5.1 Voice quality and language status

Yuasa (2010) compared the speech of young educated Japanese and American college students and observed that female Japanese speakers would use fewer occurrences of creaky voice than their American counterparts. Yuasa (2010: 132) hypothesized that "Japanese women may continue utilizing a high-pitched voice as a reflection of the persistent societal expectation to project a feminine image. [...] This follows from the fact that appearing feminine is still important in this society". This is in line with Ohara's (1999) observations that Japanese female subjects produced higher f_0 values when speaking Japanese rather than in English.

Conversely, the fact that creaky voice is now a common feature in the speech of young

American women may be due to the fact that, through years, women have intuitively realised that their socio-professional success depended on assimilation within the dominant group, thus, within men (Yuasa, 2010). Benoist-Lucy & Pillot-Loiseau (2013) analysed the usage of creaky voice of American women learning French, in both English and French, and observed more occurrences of creaky voice in English. Creaky voice is known to be part of the language setting in English, but not in French, which is frequently associated with nasality and breathiness (Esling & Wong, 1983). Benoist-Lucy & Pillot-Loiseau (2013) concluded that these speakers integrated the fact that voice quality attitudes can be language-specific, and reduced their usage of creaky voice when speaking French²³.

3.5.2 Voice quality as an index of group membership

When hearing a voice, individuals tend to spontaneously categorise the speaker as a member of one specific social group, even within a particular language or dialect group. Any individual has once formed impressions of a speaker's social status, ethnicity, and so forth²⁴.

Esling (1978) found that, in Edinburgh English, the use of creaky voice is associated with a higher social status. Similar findings were observed for Glasgow English (Stuart-Smith, 1999). Pittam (1987) had both Australian and American speakers evaluating different voice qualities produced by Australian speakers, and observed that creaky voice was mostly associated with high social status. Voice quality appears, therefore to be a marker of social class.

Voice quality also seems to index geographical origins within one country, meaning that voice quality may be varied differently across dialects too. Henton & Bladon (1988) studied the speech of 40 male and 40 female in two dialects of British English (RP and Modified Northern). They observed that male speakers of both varieties would use more instances of creaky voice than females, with male speakers of Modified Northern using more of them.

²³However, they do not exclude the possibility that this difference may have been induced by the cognitive weight of speaking another language, which may have left more vigilance for specific phonatory usage.

²⁴Stereotyping based on language, sometimes refer to as *voice stereotypy* (Aronovitch, 1976), is extremely common. Stereotypes associating certain kinds of voices with a specific ethnic group can have dramatic and harmful consequences. Purnell *et al.* (1999) observed that listeners can identify non-standard, ethnically marked dialects on the basis of very short speech segments, and that landlords may use this information in discriminating against potential tenants.

3.5.3 Voice quality and role identity

Speakers can vary voice quality for managing personae they may possess. The idea is essentially to convey different images.

Hall (1995) observed that a phone sex operator might adopt a breathy voice to take on the persona of someone who is constantly aroused. Creaky voice was also found to index the performance of a sexy persona by Japanese porn actresses (Kajino & Moon, 2011), or a woman's dangerous sexuality in Chinese TV drama (Callier, 2010). In Henton & Bladon's (1988) study mentioned in the previous section, more instances of creaky voice in the speech of male than female speakers were observed. The authors concluded that creaky voice could be regarded as a "robust marker of male speech" and was mostly used to sound "hyper masculine". In her study on gender variation in Dutch voice quality, Biemans (2000) found that a low, loud, creaky voice conveyed an image of masculinity, while the opposite characteristics were linked to femininity.

Yuasa (2010) observed that young Northern Californian women would speak with a creaky voice to index a professional, upwardly-mobile, female persona. Conversely, Anderson *et al.* (2014) found that vocal fry was negatively perceived when used by young female American speakers in a labour market context. Young adult female speakers exhibiting vocal fry were perceived by American adults as being less educated, less trustworthy, less competent, less attractive, and less hireable. These negative judgments were stronger when the listener was also a woman. Wolk *et al.* (2012) analysed the speech of young adult female speakers of standard American English (18-25) in a reading task and observed that more than 2/3 used vocal fry. A comparison made with young adult male speakers of American English in a follow-up study showed that vocal fry was four times less prevalent in the speech of male speakers (Abdelli-Beruh *et al.*, 2014). Pennock (2015) studied the speech of three American actresses playing both American and British characters to show that they deliberately manipulated voice quality to attain specific "styling" effects. He analysed and compared the number of creaky occurrences in films in which these actresses played a role that embodies positive stereotypes of femininity. Results showed that creak was heard more frequently in American acting, supporting the fact that it can be used as voluntary articulatory setting. Pennock concluded that desirability thus depends on cultural setting, for creak is considered to be desirable in America, but less so in Britain.

Creaky voice is a frequent component of the voices of American actresses who act as role models to many young women in America, which may explain that its usage enhances their desirability. This echoes Gottliebson *et al.*'s (2007) study in which college students reported that vocal fry was deliberately used by many speakers of this age group. They hypothesized that these college students have either practiced, or perceived this voice quality, and modelled it to match popular figures.

We have seen in this section that non-modal phonation types do not have the same function across language. Phonation type may be contrastive in some languages, meaning that the contrast is implemented in the language's phonology, while it can be used as an allophonic variant, or to mark prosodic boundaries in others (e.g. English). Voice quality also carries indexical information and can be varied to signal speakers' affects and emotions, or to convey different images.

4 Conclusion

Voice source variations partly arise from complex laryngeal configurations. Speakers appear to make use of different strategies available to vary voice quality. Those strategies include, among others, increasing or decreasing f_0 , manipulating glottal gap, changing OQ, and altering the skewness of glottal pulses (Kreiman *et al.*, 2012). Voice quality variations serve as a vehicle for conveying numerous types of information. Variations can be linguistically or socially meaningful. More precisely, phonation types can be manipulated to produce phonological and allophonic contrasts, or to mark prosodic boundaries, depending on language. They can also characterise different mental and affective states, or carry indexical information. Voice source variations are, therefore, predictable in certain circumstances. A speaker may also deliberately make use of a specific voice quality to convey a specific image. It then appears that voice quality settings can be learned and acquired. The process of adaptation, also known as accommodation, is the focus of the next chapter.

CHAPTER II: Phonetic convergence

Phonetic accommodation is the process by which speakers adjust their speech patterns in response to their interlocutors. This process can be split into three distinct sub-categories: divergence, maintenance and convergence. During convergence, the phonetic distance between talkers is shortened. A decrease in the dissimilarities of acoustic-phonetic forms between talkers is observed (Pardo, 2006). No variation in phonetic distance is observed during maintenance. In the case of divergence, the phonetic distance between talkers is increased. Speakers accentuate speech differences between themselves and their interlocutors. Speakers can also converge on some acoustic dimensions and diverge on others (Bilous & Krauss, 1988; Pardo *et al.*, 2012). The amount of inter-speaker variability can vary but is often fairly important. Speakers can, for instance, converge on the same set of acoustic attributes, and other speakers on a different set of attributes.

The continuously growing interest in the phenomenon of convergence has led to numerous theories of cognitive systems and social interaction to be developed. From one perspective, motivations for convergence are said to be automatic and cognitive. Convergence is seen as an unavoidable consequence of the way language is processed in the brain, and to result from an unmediated link between perception and production (Sancier & Fowler, 1997, Goldinger, 1998). From another perspective, phonetic convergence is said to be modulated by social concerns such as gender, conversational role, likeability, or cultural affiliation, for instance (Namy *et al.*, 2002; Pardo, 2006; Pardo *et al.*, 2010; Nilsenová & Swerts, 2012, Babel, 2012; Babel *et al.*, 2014). From the hybrid perspective, both perceptual-motor factors, as well as social and psychological influences, are active when convergence is observed.

CHAPTER CONTENT

This chapter provides an overview of the process of phonetic convergence, sometimes referred to as *adaptation, entrainment, imitation, or alignment*. The different linguistic levels at which convergence has been observed and popular experimental designs used to assess convergence are reviewed in the first section. In the second section we discuss the psycholinguistic and automatic approaches on phonetic accommodation. We first report theories arguing that convergence results from an unmediated link between perception and production. We then focus on cross-linguistic accommodation and on how it is tightly linked with L2 learners' ability to perceive the linguistic features present in the L2. In the third section we discuss the external and social influences that drive the process of accommodation. Convergence effects have been observed as varying greatly depending on factors such as gender, conversational role, cultural affiliation or perceived attractiveness/liking. Creaky voice imitation being the main topic of this dissertation, we conclude that chapter by reporting the very few studies conducted so far on creaky voice alignment.

1 Measuring accommodation

We briefly review the main linguistic levels at which convergence has been observed in the literature in section 1.1. We discuss the standard experimental practices in section 1.2.

1.1 Levels of measurement

Convergence has been measured at many different levels of linguistic systems. At the syntactic level, the syntactic structures used by one speaker affect the interlocutor's own use of syntactic structures (e.g. Branigan *et al.*, 2000). At the lexical level, speakers increasingly use the same lexical items as they coordinate their perspectives (e.g. Bell, 2001). Convergence in utterance length or pausal phenomena has been observed at the prosodic level (Bilous & Krauss, 1998). Speech rate, fundamental frequency (Pardo *et al.*, 2010; Sato *et al.*, 2013), or vowel duration (Zajac, 2013) are suprasegmental features speakers often converge on. Convergence can even occur in fine-grained phonetic details like in the F1-F2 vowel space (e.g. Babel, 2009, 2012; Pardo *et al.*, 2010, 2017).

1.2 Popular methodological practices

The AXB test is the most common method used to perceptually assess convergence. In this test, the degree of a participant's change in a certain acoustic dimension before and after exposure to the target speaker is compared. It is particularly convenient for it allows taking multiple acoustic dimensions into holistic consideration. This test is used in both interactive and non-interactive tasks.

Interactive tasks are designed to elicit convergence in spontaneous conversations. The interactive task found across many convergence studies is the Map Task. In this task, each of the two participants has a map which the other cannot see. The speaker designated as the Instruction Giver has a route marked on their, while the other – the Instruction Follower – has no route. The goal of the task is to reproduce the Instruction Giver's route on the Instruction's Follower's map. Successful communication is, therefore, very important. At the beginning of the session, participants are explicitly told that their maps are not identical. It is their role to discover how the two maps differ while trying to reproduce the route. This task offers several advantages. The

experimenter can control the information shared by the participants by arranging mismatches between landmarks, choosing their names and their locations on the maps. The names of the landmarks can be designed to be of phonological interest. The other advantage is that the pairing of the participants can be controlled by the experimenter to fit theoretical motivations (Anderson *et al.*, 1991).

Non-interactive tasks are used to elicit convergence in less natural and more controlled settings. Speech shadowing is the most widely used type of non-interactive tasks. In such tasks, speakers are first required to read aloud words, which will serve as the baseline production of each speaker. Those same words previously recorded by a model speaker are then presented auditorily to the speakers, who are instructed to repeat them as naturally as possible (e.g. Goldinger, 1998; Namy *et al.*, 2002; Goldinger & Azuma, 2004). Stimuli produced before and after exposure to the model talker can then be compared. The advantages of this task are the same as those of the Map Task. Variants of both tasks can be found across different studies (see, for instance, Dufour & Nguyen, 2013; Pardo *et al.*, 2018; , or Wagner *et al.*, 2021).

In this section we saw that convergence has been measured at many different linguistic levels, from syntactic structures to fine-grained phonetic details. Although we reported the most popular methodological practices used to assess convergence, it is common to find variable designs in the literature, depending on the research question the studies address²⁵. In the next section we discuss the psycholinguistic and automatist approach on phonetic accommodation.

2 The psycholinguistic approach on phonetic accommodation

We report theories supporting the automatist approach on convergence in section 2.1 and include a section on cross-linguistic accommodation in section 2.2. Models of L2 perception and acquisition, the effect of proficiency and that of 'phonetic talent' are presented in section 2.3.

2.1 Convergence as an automatic process

Some theories posit that convergence is automatic. Convergence as the result of an unmediated perception-production link can be understood through two different ways: as a result of

²⁵See Florent Chevalier (2021) for a study on accommodation using GAMMS, for instance.

exemplar-based lexical representation (Goldinger, 1998), or as a natural consequence of general processing such as priming (Sancier & Fowler, 1997; Pickering & Garrod, 2004).

In the Episodic Theory of Speech Perception and Production (Goldinger, 1998; Goldinger & Azuma, 2004), it is proposed that convergence results from the detail of episodic memory traces – referred to as echoes – which are elicited during the production phase. Each individual heard word is said to leave a trace in memory. These echoes are information about all the activated traces previously stored (e.g. characteristics of a speaker’s voice), to which will be added the most recently trace called the input. This perceptual event can influence the mental representation of a word, affecting the subsequent production of that same word. The activation of traces during perception and production can, therefore, be responsible for convergence effects. The first prediction of this model is that, in a shadowing task, low frequency words are more subject to convergence than high frequency words because they are represented by fewer traces in memory. The greater part of the activation in the echo comes from the stimulus itself and shows more convergence effect. The second prediction is that the aforementioned imitative fidelity is said to increase with increased exposure to that stimulus. The more times a listener will hear the same stimulus, the greater the number of traces that are identical to that stimulus will be stored in their long-term memory. Consequently, these traces will have a very strong influence over the shape of the echo. The third prediction is that more immediate shadowing responses display greater imitative accuracy. The more a speaker is asked to hold a word in their working memory, waiting to shadow it, the more they will create new echoes, creating a 'feedback loop', which will cause “each successive echo [to] drift toward the central tendency of all prior traces in LTM [long-term memory]” (Goldinger, 1998: 256). Consequently, the idiosyncrasies of the new stimulus will gradually be erased and the production of the shadowed word will display less imitative fidelity than that of an immediate response.

According to Sancier & Fowler (1997), there is no social motivation for convergence to arise cross-linguistically, while social affiliation might be partly responsible for convergence effects observed across dialects of a common language. They collected data from a bilingual speaker of American English and Brazilian Portuguese at three different points over a year. The first session took place after the speaker stayed in the US for 4.5 months, the second session following

a 2.5 month stay in Brazil, and the last session after a 4 month stay in the US. They compared the production of VOT for /p/ and /t/ in all 120 English and 120 Portuguese sentences collected during each session, and observed convergence effects towards the standards of the language community to which the speaker had been most recently exposed to. English phones became more Portuguese-like (e.g. VOT shortened) after recent experience to Brazil, and vice versa. Sancier & Fowler (1997: 421) concluded that speakers experience “perceptually-guided changes in speech production” that are automatic and result from exposure to a specific language community. These changes affect the phonetic realisations of phonological segments. The authors refer to this phenomenon as 'gestural drift', which is directly linked to the gestural understandings of speech perception developed in two different theories: The Motor Theory of Speech Perception (Lieberman & Mattingly, 1985; Lieberman & Whalen, 2000) and Direct Realist Theory (Fowler, 1996; Sancier & Fowler, 1997). These theories argue that perception of speech units are direct and not defined in terms of acoustic properties of the speech signal but rather in terms of articulatory gestures. A perceived gesture is said to constitute instructions that are necessary for reproducing that gesture, therefore serves as motivations for imitation. Within the framework of The Motor Theory of Speech Perception, listeners are said to recruit their speech motor system during speech perception and prime it. Automatic priming might, therefore, be the reason behind convergence effects observed cross-linguistically. Sancier & Fowler (1997) also observed that convergence may not occur completely. They relate this finding to prior experience to one or the other language, and to what they call 'recency effects' whereby more recent experience “exerts a disproportionately stronger impact on current perception and behavior than more distant past experiences” (Sancier & Fowler, 1997: 432). This parallels Goldinger’s (1998) exemplar-based theory of lexical storage and access.

Another contribution to the automatist position was provided by Pickering & Garrod (2004). They suggested that imitation serves a purpose in dialogue processing and that the most efficient way to achieve successful interactions is for interlocutors to automatically align at various levels of linguistic and situational representations, instead of modelling one’s interlocutor mental state. From this perspective, any individual relies on the same representations during both the perception and production phases. This balance of representations brings the two interlocutors’

numerous levels of representation into alignment through automatic priming. This results in the interlocutors sharing the same representation at a given level. Dialogues become less cognitively demanding as alignment increases, which is due to the reduced need to make intricate choices as to how to portray the mental state of the interlocutor. Increased mutual understanding therefore results from the alignment of speakers' representations.

Dufour and Nguyen (2013) compared the phonetic convergence effect when Southern French speakers were explicitly instructed to imitate stimuli they were exposed to, to the effect observed in a shadowing task. Bisyllabic words ending in /e/ or /ɛ/ were randomly displayed on a computer screen during the pre- and post-test. They were instructed to read the words as clearly and naturally as possible. During the test, the speakers were presented the stimuli over headphones. 10 subjects were instructed to imitate the speaker's specific pronunciation and 10 others to repeat it as clearly and naturally as possible. They observed a greater effect in the imitation group, but only during the test phase. This effect probably relied on attentional factors:

Given that participants were asked to imitate the specific pronunciation of the speaker, they have likely paid greater attention to the speaker's indexical features in order to get as close as possible to the specific pronunciations of the words they heard.

Dufour & Nguyen (2013: 5)

They observed consistency of phonetic convergence across the two different settings (shadowing and imitation) in the post-test reading task, providing further evidence that convergence emerges from an automatic perception-production integration mechanism.

2.2 The effect of language status

Cross-linguistic accommodation has been extensively studied over the years. Zajac (2013) conducted an experiment on how phonetic imitation can, or cannot, be influenced by the model talker being a native or a non-native speaker of English (i.e. Polish). She investigated the variability in duration of the English vowels /æ, e, ɪ, i:/ in both shortening and lengthening

b_t and b_d environments. The idea was to analyse “whether articulatory habits typical of the participants’ L1 would prevent them from imitating this L2 feature” (Zajac, 2013: 21). The durational characteristics of this set of vowels were chosen as pre-fortis clipping²⁶ which is “a feature characteristic of English pronunciation [that] may cause difficulties for Polish learners” (Zajac, 2013: 21). English is a vowel-length sensitive language whereas Polish is not (Waniek-Klimczak, 1998). Polish learners of English were interviewed by both a native speaker of English and a native speaker of Polish talking to them in English. Informants were found to converge towards the native model talker, and to diverge from the non-native model talker. The author suggested that subjects might have been aware of the foreign accent of the non-native speaker, leading them to diverge from her in order to distance themselves from other foreign-accented talkers, and to converge towards the native model in order to sound more native-like. Murphy (2014) studied the production of English learners of French when interacting with a native and a non-native speaker of French interacting with them in French²⁷. Similar results were obtained as those by Zajac: English learners of French converged more towards the native speaker of French than towards the non-native speaker.

Burin & Ballier (2017) analysed two spontaneous conversations between a French learner of English and a native speaker of British English (one male, one female). Convergence in vowel spectra, vowel duration, and speech rate, were observed. Interestingly, not only the participant converged to the model talkers in terms of speech rate, but the model talkers also converged towards the participant. This adaptation process towards L2 speech is often referred to as 'foreigner talk' and is defined by the use of a higher pitch, shorter sentences, and a slower speech rate (Snow, 1995). Smith (2007) found that native speakers of French would use an expanded f_0 range when talking to non-native speakers. In a recent study, Wagner *et al.* (2021) analysed convergence in a disguised memory task in which native speakers of Dutch were asked to repeat backwards a list of words of varying length produced by a female native speaker of a Serbo-Croatian dialect. Speakers were found to converge on a number of dimensions to the non-native

²⁶English vowels are subject to pre-fortis clipping when they are followed by a fortis consonant within the same syllable. Articulation is said to be strong and energetic. They tend to be long, voiceless, aspirated and high (Collinset *al.*, 2019).

²⁷Similar conditions as in Zajac’s study, though the focus of Murphy’s was formant frequency dispersion and not vowel duration.

speaker (e.g. speech rate, vowel duration, f_0 , etc.). However, the stronger the non-native accent was rated by the participants, the less likely they were to converge.

These different studies provide strong evidence that non-native speakers use identical accommodation strategies than native speakers. As we saw in a previous section, convergence results from an unmediated link between perception and production. Convergence from L2 learners therefore results from their abilities to perceive L2 linguistic features and to be able to reproduce them.

2.3 Perception of L2 sounds

2.3.1 Models of L2 perception and acquisition

There is a great deal of variation in how well and fast adult learners acquire the sound system of a second language. Numerous causes can account for those differences. In views on Foreign Accent, external circumstances like age of arrival or length of residence in a foreign country, age of learning, and/or the amount of exposure to the L2, are said to account for such differences (Flege *et al.*, 1995; Piske *et al.*, 2001; Birdsong, 2006; Best & Tyler, 2007). The acquisition of L2 pronunciation is tightly linked to the perception of L2 contrasts which depends on how relevant sounds are distributed in the L1 and the L2. L2 learners tend to have difficulty differentiating two L2 sounds that do not contrast in the L1, especially when they are phonetically similar to an L1 sound²⁸. Several models have been proposed to explain this difficulty in perceiving L2 sounds. They are briefly reviewed in the following:

1. **Perceptual Assimilation Model – PAM** (e.g. Best, 1995; Best & Tyler, 2007): L2 sounds are perceived according to their similarities to L1 sounds that are closest in terms of articulation. The ability to discriminate a non-native contrast depends on how each of the contrasting sounds is assimilated to L1 sounds. Several assimilation patterns can occur among which the single-category assimilation which asserts that, if X_{L2} and Y_{L2} are equally similar exemplars of a single L1 sound, then discrimination will be very poor. Conversely,

²⁸e.g. /i/-/ɪ/ in English which are phonetically similar to /i/ in French.

the non-assimilable pattern claims that, if both sounds fall outside the L1 phonetic space, then discrimination will be very good.

2. **Speech Learning Model – SLM** (e.g. Flege, 1995; Flege & Bohn, 2021): whether new phonetic categories for L2 sounds can be formed with increasing experience with the L2 depends on the perceived distance between the L2 sound and the closest L1 sound, but also on the age of learning. If two sounds are identical in the L1 and the L2, L1 category will be used (equivalence classification²⁹). If two sounds are similar, but the phonetic difference is undetectable, no new phonetic category will be formed (equivalence classification). If an L2 sound is not present in the L1, a new phonetic category will be formed, but with experience.
3. **Native Language Magnet Theory – NLM** (Iverson & Kuhl, 1995): having a native language ultimately affects the way we perceive L2 sounds. L1 sounds develop phonetic prototypes that act as magnets with reference to other sounds. The L1 warps the perceptual space, resulting in perceptual difficulties in cases where L2 sounds resemble those prototypes.

2.3.2 The effect of proficiency

Although foreign accent is often subject to a phenomenon called 'fossilization', which is defined as a permanent non-native state appearing during learning (Murphy, 2014), the production of proficient speakers has been found to be both acoustically and perceptually similar to the output of native speakers. Rojczyk (2013) analysed the speech of Polish learners of English and observed that their production of the English vowel /æ/ was closer to that of the model talker as a result of exposure. Swerts & Zerbian (2010) compared the prosody of proficient vs. less proficient speakers of Black South African English having L1 as their mother tongue. The production of proficient speakers was found to be equivalent to the output of native speakers of Black South African English in their use of intonation for marking focus and boundaries. The less proficient speakers would mark boundaries in a similar way as L1 speakers of Black South

²⁹Equivalence classification is “a basic cognitive mechanism which allows humans to perceive constant categories in the face of the inherent sensory variability found in the many physical exemplars which may instantiate a category.” (Flege 1987: 49).

African English, but followed the rules of their native language for signalling focus (transfer effects from the L1 to the L2). Wang & Gu (2022) analysed the effect of language proficiency on phonetic accommodation in Chinese EFL learners. High-proficiency learners were found to converge more than low-proficiency ones in terms of vowel duration. The literature therefore points in the direction that more experienced listeners (listeners that have received more input) are better at categorizing and discriminating certain non-native L2 contrasts significantly better than listeners that have received less input (Best & Tyler, 2007). As a result, high proficient learners may be better imitators than low proficient ones.

2.3.3 The effect of 'phonetic talent'

Other causes that address learners' individual abilities and characteristics (i.e. intelligence or personality) can impact the acquisition of L2 contrasts. Some assume that the pronunciation learning abilities are inherent to the speaker. This is what Lewandowski & Jilka (2019) call 'phonetic aptitude' or 'phonetic talent'. As we previously saw, successfully acquiring a given L2 sound system requires the ability to accurately perceive the phonetic details of that language, and the ability to correctly reproduce these characteristics. Accommodation being partly the result of an internal automatic perception-production link, learners that are considered good at acquiring the phonetic features of an L2, or of a given dialect, might also be very good at aligning to their conversational partners. Particularly, a learner that already has a near-native accent might also be a great converger (Lewandowski & Jilka, 2019). Lewandowski & Jilka (2019) analysed phonetic convergence in an L2 dialogue between German L2 learners of English and native speakers of English. Based on test results and evaluation in a series of extensive language tests, learners were divided into two groups: phonetically talented and less talented. Results showed that a specific phonetic talent component increases convergence. Phonetic convergence is, therefore, also susceptible to internal factors such as executive attention and working memory components, observations already made by Dufour & Nguyen (2013). Such individual differences might account for the great deal of variability often observed in studies on phonetic accommodation (see Burin, 2018, who observed no proficiency effect, for instance).

In this section we discussed theories arguing that accommodation is automatic and results

from an unmediated link between perception and production. We reported findings on cross-linguistic accommodation and on how it is linked with the ability that L2 learners have to perceived the different linguistic features present in the L2. We addressed the effect of both proficiency and 'phonetic talent' on accommodation and acquisition of L2 speech features. In the next section we discuss the sociolinguistic approach on phonetic accommodation.

3 The sociolinguistic approach on phonetic accommodation

Convergence is also said to be driven by external and social influences. We review the effect of gender in section 3.1, the effect of interactional dominance in section 3.2, the effect of cultural affiliation in section 3.3, and the effect of social preference in section 3.4.

3.1 The effect of gender

Gender effects have been extensively studied in research on accommodation and women were more often found to converge more than men. Pardo (2006: 2388) argued that the tendency for female speakers to converge more than male speakers can be due the fact that “women might be more sensitive to indexical features³⁰ of talkers, which could lead to greater convergence”. Babel (2012) observed more convergence from female speakers than from male speakers in a shadowing task. Eisikovits (1987) analysed the speech of teenagers (intergroup girls/intergroup boys) and discovered that female informants converged more towards her, the female interviewer, than male informants, who even tended to diverge from her speech. Namy *et al.* (2002) conducted an experiment to assess whether gender differences in vocal accommodation occurred in socially minimal situations. Participants (both male and female) were instructed to repeat isolated words produced by different talkers. Female speakers converged more than male speakers, and both groups converged more to male than to female speakers. In the literature, it has often been reported that male and female speakers have a tendency to converge more towards male interlocutors (e.g. Bilous & Krauss, 1988; Gallois & Callan 1988; Willemyns *et al.* 1997). The effect of gender on convergence is, however, inconsistent across studies. Pardo (2006) compared

³⁰Indexical features are information about a person such as physical, social, and psychological characteristics (e.g. age, gender, social status, and emotional state).

convergence effects in an interactive task and observed more convergence from male than from female speakers. In her study on accommodation between French learners of English and native speakers of British English, Burin (2018) observed more convergence from male than from female speakers.

Gender seems to interact with multiple factors in phonetic convergence, which accounts for the fact that different settings lead to conflicting results. Wang & Gu (2022) analysed the effects of gender and language proficiency on convergence in vowel spectra and observed less convergence effects in female speakers than in male speakers in the advanced learners' group, while the reversed was observed in the lower-proficiency learners' group. Pardo *et al.*, (2017), observed overall convergence on AXB perceptual assessment (listeners were asked to judge whether shadowed items were more similar to the model items than baseline items) and duration, minimal convergence in the F1-F2 vowel space and F2 alone, and no convergence effect in F1 and f_0 measures. No effect of gender or word frequency alone were observed but female shadowers were found to converge more to low frequency words, and more than male shadowers to low-frequency words. The authors concluded that female speakers seemed to be more sensitive to lexical features than men, and that the reason why female speakers were found to converge more in some studies might be due to the use of low-frequency words only (e.g. Namy *et al.*, 2002; Babel *et al.*, 2014). Similarly, studies in which low-frequency words displayed greater convergence effects only used female speakers (Babel, 2010; Dias & Rosenblum, 2016).

3.2 The effect of interactional dominance

It is often said that, in dyadic conversations, each person is assigned a role and speakers converge more towards the person who is more socially dominant, who has more 'power' (Watzlawick *et al.*, 1967; Nilsson & Swerts, 2012). Giles (1973) analysed the conversation between an inspector and a traveller on a train and observed more convergence effects from the latter towards the speech of the former. In Polish student-teacher conversations, Andreeva *et al.* (2021) observed overall convergence from the students towards the teacher. Additionally, any status difference might lead to an asymmetry in accommodation: a person with lower status should accommodate to the person with higher status (Bulatov, 2009). Pardo (2006) conducted an

experiment using a Map Task to analyse the effect of conversational role on convergence effects. The idea was to determine which of the instruction giver or the instruction follower would converge more towards the other. Instruction followers were expected to converge more towards instruction givers than instruction givers towards instruction followers. Overall conclusions did not, however, follow these predictions and went in the opposite direction. This can be explained by the fact that gender was also found to interact with talker role, indicating that interactional dominance might be affected by other social factors.

3.3 The effect of cultural affiliation

In addition to gender and talker role, both cultural affiliation and language attitude can influence convergence. Giles & Johnson (1987) found evidence that a non-native speaker is likely to imitate a native speaker if both share significant social identities, related to ethnicity or not. Zuengler (1982) demonstrated that L2 pronunciation can vary, by diverging or converging, if a native English-speaking interlocutor conveys negative or positive attitude towards the ethnic group the non-native speaker belongs to (here L1 Spanish or L1 Greek speakers). In her study, non-native speakers who perceived threat would phonetically diverge if they identified firmly as ethnic group members, or if they wanted to defend their ethnic solidarity. Babel (2010) conducted a shadowing task in which she attempted New Zealand English speakers to accommodate to Australian English by making them believe the Australian model talker had either positive (flattering) or negative (insulting) views of New Zealand. Participants were also required to complete an Implicit Association Task (IAT; see Greenwald *et al.*, 1998) to evaluate how strongly they were pro-Australia or pro-New-Zealand. Significant convergence effects in vowel spectra were observed in both the shadowing and post-shadowing task, meaning that they retained the same convergent pattern beyond the shadowing context. Each of the vowels, word frequency, and IAT scores significantly displayed convergence for the shadowed items. More specifically, subjects appear to have converged more on some of the vowels, on lower-frequency words, and if they held a pro-Australia bias. For the post-shadowing task, only IAT scores significantly predicted convergence, meaning that a pro-Australian bias influenced the likeliness of subjects to retaining convergent vowels. The positive/negative variable failed to reach significance in both of these

tasks. These subjects may have attempted to decrease social distance between them and the model talker, or to affiliate themselves with Australians as a group through convergence.

3.4 The effect of social preference

Previous findings show that speakers tend to converge towards an interlocutor they appreciate and from whom they want to be appreciated (Giles *et al.*, 1991; Byrne, 1997; Chartrand & Bargh, 1999; Babel, 2009). Speakers who are mimicked by their conversational partners will, in turn, like them more, resulting in a higher degree of harmony in the interaction (Nilsenová & Swerts, 2012). A somewhat related effect on phonetic accommodation is that of attractiveness. According to the similarity attraction hypothesis (Byrne, 1971; Bourhis & Giles, 1977), people tend to act more like those they are attracted to. If we relate that claim to speech, it means that speakers first need to perceive their interlocutor's speech, assess their social attractiveness, and finally opt for an accommodative strategy. The assessment of social liking can also lead to divergence in order to accentuate differences and increase social distancing with one's interlocutor. Babel (2012) conducted an experiment in which informants (both male and female) were randomly assigned a model talker. They were separated in two different conditions: one in which a picture of their assigned talker was displayed, and the other in which they would only hear the voice of the model talker. The idea was to determine whether accommodation resulted from liking measured here through attractiveness. Results showed that informants accommodated more in the visual condition. Patterns of accommodation were, however, found to be different for male and female speakers. Female speakers converged more towards the model speaker they rated as attractive, whereas the more attractive the male speakers rated the model speaker, the less likely they were to converge towards him. Again, it seems that different social factors interact with each other.

In this section we saw that different external and social factors can drive the process of accommodation. We reported studies in which gender, interactional dominance, cultural affiliation or social preference were found to have an effect on convergence. In the next section we report the very few studies that have been conducted on creaky voice accommodation.

4 Creaky voice accommodation

Women have been found to produce consistently more occurrences of creaky voice over the years. Originally, this phenomenon was essentially observed in the speech of men. Men speaking with a creaky voice were perceived as being more authoritarian, intelligent, and self-confident (Esling, 1978; Yuasa, 2010). It has been hypothesized that women are producing more instances of creaky voice – therefore to converge towards men – in an attempt to sound like them and to convey the same image (Pennock, 2015).

Voice quality accommodation is an area that remains to be widely explored. To our knowledge, only two studies are concerned with convergence in creaky voice in interactive designs. Borrie & Delfino (2017) analysed conversational entrainment of vocal fry in young adult female American English speakers and found that they would use more vocal fry when interacting with a partner who exhibited substantial vocal fry. Subjects who displayed greater convergence effects towards their conversational partner also reported more enjoyment of the conversational experience. These observations are in line with the idea that aligning with a person’s behaviour results in increased harmony in the interaction, as we previously saw (Nilsenová & Swerts, 2012). The authors concluded that “although sociocultural motives play some role in driving the prevalence of vocal fry in conversational speech behaviors, [...] the use of vocal fry in spoken dialogue involving young American women is modulated by the pervasive behavioral matching phenomenon of conversational entrainment” (Borrie & Delfino, 2017: 30). Pillot-Loiseau *et al.* (2019) analysed the evolution of creaky voice usage in a collaborative reading task between native speakers of French and native speakers of English. Subjects were separated in 9 tandem pairs made up of one native speaker of French and another one of English. Each tandem pair met 12 times on average over a 3-month span. Semi-spontaneous narrative and a monitored reading task were recorded every time in both languages. This study only focuses in the collaborative reading task in which the two partners were asked to read a text in their respective L2s first, and then in their L1s. They observed a positive correlation between the L1 and the L2 speech of the same speaker, indicating that creak usage developed in the L2, which might result from accommodation to their tandem partner. The authors argued that this correlation points to either transfer of creaky voice from L1 to L2 (native English speakers), or reverse transfer (L2 to L1; native French speakers)

after interacting over a three month span. They concluded that the habits developed in the L2, here creaky voice usage, might have influence its usage in the L1.

5 Conclusion

In this chapter we saw that convergence is assumed to result from an internal automatic perception-production link and is influenced by both external social factors and communicative goals to express the kind of relationship existing between interlocutors (Coles-Harris, 2017). Convergence is also sensitive to phonetics and phonology. In the shadowing experiment conducted by Babel (2012), briefly mentioned in the section on gender effect, low vowels displayed greater convergence effects than high vowels. As claimed by the author, low vowels tend to exhibit more variation in North American English and most of the subjects were from California, some from the Upper Midwest. The Midwestern speakers were found to be the best convergers, and more convergence was observed on the /a/ vowel, which was the vowel that displayed more differences as compared to that of the model talker. She hypothesized that a greater distance in the phonetic space might allow for greater convergence effects. More convergence on low vowels, and especially in the F1 dimension, was also observed in Burin & Ballier (2017).

Although we tried to categorize the different effects observed in the literature, different socio-contextual and linguistic factors, taken altogether, influence phonetic convergence. One factor alone cannot account for the asymmetrical behaviours and inconsistencies observed across studies. For instance, females were found to converge more to low than to high frequency words in the shadowing task conducted by Pardo *et al.* (2018). No difference was observed for male speakers, indicating that women might be more sensitive to lexical factors than men. Female speakers were also found to be more influenced by task settings and/or talker identity in that same study. The authors also argue that it might be possible that speakers with similar phonetic attributes converge more in the same dimensions than others. Differences in experimental designs, talker-related variability, external social factors, and idiosyncratic patterns of convergence, make it difficult to draw conclusions and interpretations in accommodation studies.

In the next chapter we define our experimental design and the theoretical motivations of this dissertation.

CHAPTER III: Materials and Methods

In some cases researchers may recycle and analyse existing corpora to answer their research question. There existed no corpus that included the necessary material to answer our research question, hence we had to build our own. Building a corpus in speech studies means first designing appropriate stimuli and speech tasks. Participants fitting your criteria then have to be recruited and the resulting data has to be processed before further analysis. It is a tedious work that can end up being very challenging, especially due to the extreme inter- and intra-individual variability that is found in human speech. Various difficulties and obstacles may be encountered, and decisions will have to be made to cope.

This study was approved by the institutional ethics committee right before the beginning of the pandemic of Covid-19 (see Appendix D for IRB requirements). Data collection had to be suspended for a while. A very strict sanitary protocol was then set up and anyone wishing to conduct experiments had to respect it (e.g. the wearing of a mask and gloves for the experimenter, the ventilation of the booth and the disinfection of all surfaces before and after each recording session, etc). This corpus will be made available upon request to anyone who is interested in phonetic accommodation/imitation between French learners of English and native speakers of American English.

CHAPTER CONTENT

This chapter is organised as follows. We provide a description of the recording procedures in the first section. More particularly, we defined what electroglottography is, how it works, and the reasons why this technique became popular in phonation studies. The second section focuses on speech stimuli. We describe how they were designed and elicited. The recruitment process of the participants is discussed in the third section. In the fourth section we describe the whole procedure of the experiment. We discuss the language proficiency test we used to assess the participants' proficiency level and describe the different speech tasks we designed. Based on previous findings, we state our different hypotheses in the fifth section. Finally, we review the different acoustic and electroglottographic measurements we made in the last section of this chapter. We will provide information as to what software was used to extract them.

1 Recording procedures

Audio and electroglottographic (EGG) waveforms were simultaneously recorded in Praat, digitized at a sampling rate of 44.1 kHz and stored as a 16-bit wav file. The EGG signal was recorded using the EG2-PCX and the audio signal using the M80 omnidirectional headset microphone from *Glottal Enterprises*. Recordings took place both at The University of Washington (Seattle, WA) and Université Paris Cité in a soundproof chamber. The microphone was put at approximately 2 cm from the speaker’s mouth. The two electrodes were held stable by holding a velcro strap around the subject’s neck, approximately over the thyroid cartilages, at a level that approximates the position of the vocal folds. For best performance, electrode gel was applied to “help produce and maintain a low resistance between the electrodes and the skin, [because] all human skin possesses some electrical resistance and its magnitude may vary considerably among different subjects” (Colton & Conture, 1990: 15). A test of the signal was obtained by pronouncing a sustained vowel until it was confirmed that the location of the electrodes was adequate³¹ (Avelino, 2010). Speakers were also asked to record a few words before completing the task for the experimenter to make sure that none of the two signals (audio and EGG) was saturating.

We give a brief review of what electroglottography is in section 1.1 and detail how the electroglottograph is made of in section 1.2. We explain how to read an electroglottographic signal in section 1.3.

1.1 Electroglottography

Electroglottography is a non-invasive electrophysiological technique that allows the observation of the properties of the vibrating vocal folds during phonation (Avelino, 2010: 275). It was developed at the end of the 1950s by Philippe Fabre and has been largely used since the 1980s, first by researchers interested in the function of the normal larynx (e.g. Chevrie-Muller, 1967; Rothenberg, 1979; Kelman, 1981), then as a means of diagnosing vocal pathology (e.g. Smith & Childers, 1983; Childers *et al.*, 1984) as well as a tool in voice therapy (e.g. Fourcin & Abberton, 1976; Abberton *et al.*, 1977). It gradually became extremely popular among voice specialists

³¹A signal displayed on the generator indicates whether the position of the electrodes is correct or incorrect (i.e. whether it is too low or too high).

for it generates a signal that is free of supraglottal influence and is, therefore, very useful in disambiguating vocal from articulatory phenomena.

1.2 The electroglottograph



Figure 4: The EGG setup.

The electroglottograph, as can be seen in Figure 4, is made of one generator which delivers a high frequency alternating current, two electrodes positioned on the subject's neck over the thyroid cartilages, and one electronic circuit functioning as a frequency demodulator (Henrich, 2001). The EGG can be compared to an ohmmeter in the sense that it measures the voltage difference V that exists between the two electrodes. The latter is correlated with the electrical impedance of the conductor R , namely the neck, which is the resistance delivered as the current passes through the skin, the thyroid cartilage, the tissues, the muscles, and the glottis. The relationship between V and R is given by Ohm's Law $V = RI$ where V is the voltage, I the current, and R the resistance. If a steady current passes through a structure whose resistance is increasing and decreasing, then the voltage across the structure will, in turn, increase and

decrease (Baken, 1992). However, if the voltage is kept constant, then the current will decrease as resistance increases (Baken, 1992).

The larynx is an extremely mobile articulator. The cartilages shaping this structure can change their orientation in relation to each other, leading to the entire structure to be raised, lowered or tilted. These different actions will alter the relationship between the larynx and the electrodes. As a result, the location of the vocal folds in the electrical path is changed. The EGG has been designed to minimize these problems, such as minimizing the effect of changes in perilaryngeal structures, ignore or compensate for the motion of the larynx with respect to the electrodes, and compensate for changes of overall impedance (Baken, 1992, Baken & Orlikoff, 2000). If interference components remain too apparent, a high-pass filter can be applied (Colton & Conture, 1990).

1.3 The EGG signal

The current variation observed on the EGG signal is caused by “the difference in the electrical impedance of the tissues when the glottis is opened or closed and, thus, corresponds to the fundamental frequency of phonation” (Askenfelt *et al.* 1980: 258). Specifically, the electrical impedance delivered by the generator is modulated by the vibratory movement of the vocal folds (Henrich, 2001). A decreased impedance coincides with the closing phase of the glottis while an increased impedance coincides with the opening phase of the glottis (Childers *et al.*, 1984). This is due to the fact that air is an extremely poor conductor as compared to human tissues (Henrich 2001). In the course of phonation, the vocal folds are periodically separated by an air-filled space referred to as the *glottis*. During the glottal cycle, the electric impedance across the larynx falls as the vocal folds come into increasing contact, and rises as the glottis open (Baken, 1992; Baken & Orlikoff, 2000).

Another – and widely used – convention used to describe the EGG signal is by measuring the degree of contact between the vocal folds (open/closed), as shown in Figure 5. The current flow between the electrodes, therefore the signal, increases as a function of a greater vocal fold contact (Avelino, 2010). Its representation is, then, reversed compared to the one we have just seen: the signal increases during the closing phase of the glottis and decreases during the opening

phase of the glottis (Henrich, 2001).

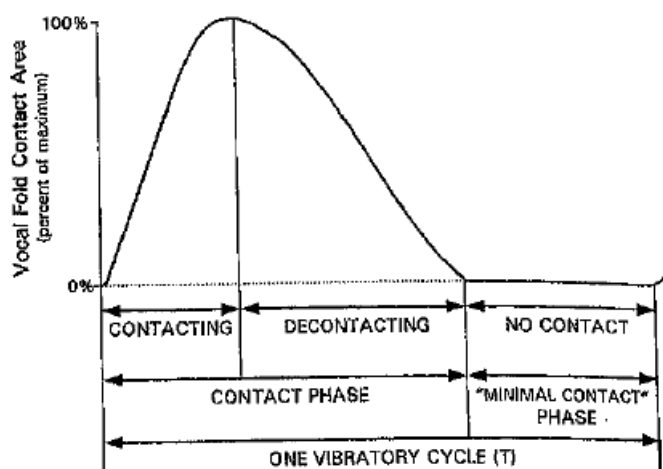


Figure 5: Divisions of the vibratory or vocal-fold contact cycle (from: Baken & Orlikoff, 2000).

The vocal folds do not meet all at once along their full length, it is a gradual process which is often referred to a “zippering” with closure starting at the anterior end of the glottis, and the closure spreading horizontally until the glottis is completely closed (Baken & Orlikoff, 2000). This can be observed on the EGG signal:

[...] after the glottis has been closed (by meeting of the lower margins of the vocal folds), contact area – and hence the magnitude of the EGG signal – continues to increase. But soon the lower margins begin to separate, and the contact area – and the EGG amplitude – decreases, although the glottis remains closed above the area of growing separation.

Baken & Orlikoff (2000: 420)

The EGG signal corresponding to one vocal-fold vibratory cycle can be described as having four main phases (Henrich *et al.*, 2004), as illustrated in Figure 6.

- **1-3) Closing phase:** contact is initiated along the lower margins of the vocal folds (1 to 2), then propagates to the upper margins (2 to 3). The closing phase occurs generally more fastly than the opening phase, resulting in a steeper slope in the EGG signal than the

one corresponding to the glottal opening (Kelman, 1981). The instant of maximum slope – which coincides with the instant of glottal closure – can be found at 2. It corresponds to a sharp positive peak in the derivative of the EGG signal. In terms of physiology, this corresponds to the instant when the lower margins of the vocal folds are completely approximated.

- **3-4) Closed phase (complete closure):** the vocal folds are in full contact, preventing air from passing through the glottis.
- **4-6) Opening phase:** the lower margins of the vocal folds start to separate gradually (4 to 5), followed by separation along the upper margins (5 to 6). The instant of maximum slope – which coincides with the instant of glottal opening – can be found at 5. It corresponds to a sharp negative peak in the derivative of the EGG signal. In terms of physiology, this coincides with the instant when the upper margins of the vocal folds start to separate.
- **6-1) Open phase:** the vocal folds are completely apart. The electrical impedance does not vary much, resulting in a relatively flat signal.

The duration that separates two glottal closure instants is more accurately determined on the derivative of EGG signal (DEGG), which makes it more commonly used. This duration coincides with the fundamental period (T_0) from which the fundamental frequency can be obtained ($f_0 = 1/T_0$). It is also easier to measure the duration between one glottal opening and the following glottal closing instant. As explained by Henrich (2001: 94), it corresponds to the open quotient OQ from which you can obtain a measurement of the closed quotient $CQ = 1 - OQ$ (Henrich, 2001).

In this section we explained what electroglottography was and how its output can bring useful additional information. Electroglottographic measurements extracted to answer our research question will further be detailed in the last section of this chapter. In the next section, we describe how speech stimuli used in our experiment were elicited.

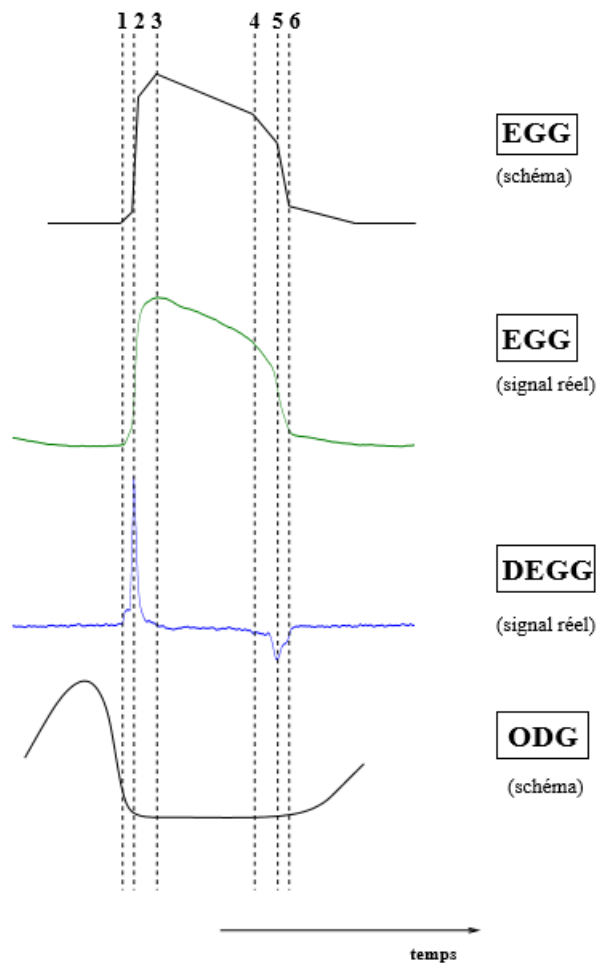


Figure 6: Figure depicting a glottal cycle as observed on an schematic and real EGG signal and its derivative (DEGG). (1)-(3): closing phase, (3)-(4): closed phase, (4)-(6): opening phase, (6)-(1): open phase. A schematic representation of the corresponding glottal flow (ODG) has been added as a comparison (from: Henrich, 2001).

2 Speech stimuli

In this section we describe how speech stimuli were designed and elicited. The model speakers that recorded the stimuli are presented in section 2.1. Instructions they were asked to follow are given in section 2.2. In section 2.3 we describe the linguistic structure of the sentences models were presented with. We explain how each file was segmented in section 2.4 and how we selected our stimuli in section 2.5.

2.1 Model speakers

Speech stimuli were elicited from 12 young, healthy male and female speakers, aged 23-32 years old (mean [M] = 26.5; standard deviation [SD] = 2.75), both at the University of Washington (Seattle, WA) and Université Paris Cité (3 French female, 3 French male, 3 American female, 3 American male)³². They were all volunteers and recruited on the basis of meeting the following:

- being a native speaker (of either French or American English)
- being aged 18-35

None of the speakers reported any history of speech or hearing disorder (Appendix C.1).

2.2 Instructions

Speakers filled out a background questionnaire and were provided with a sheet of paper on which the following instructions were given:

- read as naturally as possible (casually), at your normal pitch and rate
- repeat the same sentence 5 times in a row
- make a pause after the 5th repetition to catch your breath
- if you hesitate/stammer, restart the whole cycle of repetition of the sentence

The complete metadata questionnaire and the reported answers can be found in Appendix C.1. Speakers were then brought in a soundproof chamber and seated in front of the list of sentences written down on a sheet of paper. To make sure the instructions were clear, the experimenter gave an example orally, as if she was recording one of the sentences. Speakers were allowed to take some time to become familiar with the sentences. They had been instructed to press the *Record* button in Praat when ready, and the *Stop* button when they had successfully completed the task.

2.3 Recorded sentences

27 sentences were originally presented to the speakers (the complete list can be found in Appendix A.1 for English, and in A.2 for French). The duration of the sentences was controlled

³²28 speakers were initially recorded (6 American male, 7 American female, 8 French male, 7 French female) but most were excluded for they would not produce enough occurrences of creaky voice.

to be relatively short (5 to 7 syllables) for the participants to be able to memorise and repeat them without any difficulty during the repetition/imitation task. Only declarative sentences were used for they elicit a low and falling tone in both languages (Hirst & Di Cristo, 1998; Ladefoged, 2006). We focused on the last accented word of the prosodic unit in order to have a balanced sample between the two languages. Indeed, stress placement rules in French and English differ greatly. In English, stress can be used “to give emphasis to a word or to contrast one word with another, [or] to indicate the syntactic and grammatical relationships between words or parts of words” (Ladefoged, 2006: 110/111). Phonological stress does not exist in French and only the last syllable of the prosodic unit is accented (Fouché, 1965)³³.

To observe any possible variations due to vowel height, the last accented word is a monosyllabic word containing either a low or a high vowel (at least 10 each). According to the Intrinsic Fundamental Frequency (if_0) theory, low vowels have a lower pitch than high vowels (Whalen & Levitt, 1995). The tongue position required in high vowels pulls on the larynx, which increases the tension on the vocal folds, resulting in a higher f_0 . Since some sub-types of creaky voice are produced with low longitudinal tension of the vocal folds, it is harder to achieve on high vowels (Panfili, 2015).

The phonological pattern of each studied monosyllabic word is the following:

$$/C_1VC_2/ \text{ or } /sC_3VC_2/$$

where C_1 is either a nasal, a voiced stop, or a voiced fricative (in order to ensure a short VOT)³⁴;

where C_2 is either a voiced stop or a voiced fricative (in order to avoid voicelessness, inherent/constant creakiness, or glottal reinforcement/replacement that may have been implicitly learned);

where C_3 is a voiceless consonant³⁵.

As previously indicated, speakers were asked to repeat each sentence 5 times in a row. It was meant to maximise the chances of obtaining the right number of stimuli per speaker set to 10 same sentences produced in both non-creaky and creaky voice. Having the same sentence produced

³³However, *anaccent d'emphase* or *d'intensité* can also be present in the prosodic unit (Di Cristo, 1999).

³⁴We purposely excluded voiceless stop consonants in this position because the measurement taken at 25% of the following vowel may fall in the aspiration part of the consonant.

³⁵The structure $/sC_3VC_2/$ was used in order to get more options as monosyllabic words.

in both qualities by one speaker was crucial to avoid any sentence or speaker effect during the repetition/imitation and rating tasks. The last repetition was expected to be produced in creaky voice, as compared to the first repetition, which might be due to the low amount of air left in the lungs and/or the induced low tone.

2.4 Segmentation

Audio recordings were manually segmented at different levels using Praat. An example is given in Figure 7. Only the last word of the prosodic units was segmented at the word and phoneme level. Following Garellek & Keating (2011: 190), “the vowel onset was set at the first glottal pulse following the onset, and the vowel end was set at the last glottal pulse”.

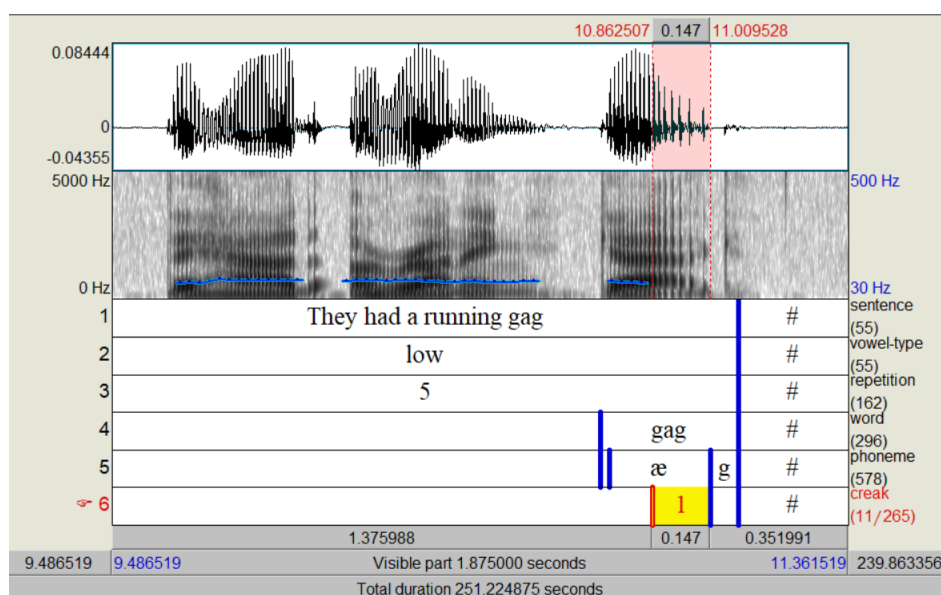


Figure 7: Example displaying file segmentation for a model speaker.

Perceptual analysis and established auditory criteria were used to identify creaky voice. Voice quality is an auditory-perceptual phenomenon (Gerratt & Kreiman, 2001) and listeners have been observed to perceptually detect the presence of creaky voice without much difficulty and with a fairly high degree of accuracy (Blomgren *et al.*, 1998). The author and another phonetically-trained rater familiar with the analysis of creaky voice independently annotated all creaky voice occurrences produced by model speakers they would observe. Each repetition was first listened,

then each word, then each vowel. The audio signal and spectrogram were checked when a portion was perceived as being creaky. If visible cues such as well-defined vertical striations, a sudden drop and/or total disappearance of the f_0 , a complete damping of pulses, an aperiodic signal, or double pulsing (Keating & Garellek, 2015) were combined with the auditory percept of creakiness, then the given portion of the vowel was annotated as being creaky. However, we saw that not all types of creaky voice have a low f_0 or an aperiodic signal, for instance. For such complex cases, only auditory cues were used to annotate creaky voice. It was extremely important for the annotators to become familiar with the speaker’s voice beforehand and to listen to it several times to remove any doubt. As observed by Epstein (2002: 43), “One person’s end-of-sentence creaky may be another person’s normal voice quality”. Boundary markers were used to mark the beginning and end of each portion of creaky voice labelled as 1. On 100% of the model speakers’ dataset, the inter-rater agreement was deemed to be really good (Cohen’s Kappa score $\kappa = 0.85$). Only sentences on which both annotators agreed were selected.

2.5 Selected stimuli

The speech stimuli elicitation procedure resulted in the collection of 232 sentence readings: 10 sentences produced in both qualities by ten speakers, and 8 sentences produced in both qualities by two speakers. Our initial objective was to use the same 10 sentences produced in both qualities for all speakers, but disagreements in annotation and great inter-individual variability forced us to use different sentences among speakers, and to reduce the number of sentences to 8 for two speakers. More specifically, we were able to only retain 8 sentences for one male French model and for one male American model, and 10 sentences for one other French male model and for one other male American model³⁶. Those sentences were selected and subsequent stimuli selection was based on them. We did our best to select sentences produced at least by 2 different speakers despite the great inter-individual variability, but the 10 sentences selected for each model speaker

³⁶For the two American speakers that produced only 8 or 10 sentences in both qualities out of the 27 presented sentences, all 5 repetitions of the 17 or 19 remaining sentences were produced in a creaky voice. The same scenario was observed in the production of the French speaker who produced only 8 sentences in both qualities out of the 27 presented sentences. Female productions were much more varied, in both languages. Creaky voice seems to be more consistently produced among male speakers, and more particularly among the male American speakers we recorded.

may differ³⁷. For the exact same reasons, it was impossible to select, for each model speaker, 5 sentences whose last prosodic unit contained a high vowel, and 5 sentences whose last prosodic unit contained a low vowel. The number of low and high vowels hence differs across language and speakers³⁸. It has to be noted that for sentences whose 5 repetitions were produced in a creaky voice, the majority of them ended with a word containing a low vowel (65.6%), which seems to corroborate the fact that low vowels are more likely to be creaky than high vowels (Panfili, 2015). It was more frequent to observe all 5 repetitions of one sentence with word-final creak in the recordings of American models (59.4%) as compared to those of French models (40.6%), more in the recordings of male American models (72.2%) as compared to those of female American models (27.8%), and more in the recordings of male French models (57.7%) as compared to those of female French speakers (42.3%).

The whole process of eliciting stimuli used in our experiment was described in this section. Several difficulties were encountered along the way, but we did our best to cope with them in the fairest possible way, not to impair too much with our initial research question. The recruitment process of the participants to which those stimuli were presented is detailed in the following section.

3 Participants

20 cisgender female native speakers of French, aged 20-30 ($[M] = 22.1$; $[SD] = 1.17$), majoring in English at Université Paris Cité participated in the study. We recruited subjects directly from classrooms, explaining we were looking for female subjects to take part in a linguistic experiment. Students willing to participate in the study were later asked to come fill out a language background questionnaire and take a language proficiency test (detailed in the following section). The aim was to make sure all of them had an advanced level of English (B2 or more according to the Common European Framework Scale of References for Languages) to avoid too

³⁷Only one selected sentence was produced by one speaker only. We had no other option than to keep it for it was produced by the speaker who had only produced 10 sentences in both qualities.

³⁸One should not think that gender is more balanced than languages in this distribution of high and low vowels: English = $N_{High} = 49$; $N_{Low} = 9$ vs. French = $N_{High} = 22$; $N_{Low} = 36$; Male = $N_{High} = 35$; $N_{Low} = 21$ vs. Female = $N_{High} = 36$; $N_{Low} = 24$.

much inter-individual variability. All participants that obtained a B2 or higher proficiency score were enabled to take part in the study. None of the participants reported any history of speech or hearing disorder.

In this section we discussed the recruitment process of the participants. The proficiency level test we used and the three different tasks (reading task, imitation task, judgment rating task) we designed are detailed in the following section.

4 Procedure

In this section we detail the procedure of the experiment. The language test we used to assess participants' proficiency level is described in section 4.1. The three different tasks in which subjects took part are presented in section 4.2.

4.1 Language proficiency test

DIALANG software³⁹ was used to assess the participants' English level. It can be downloaded free of charge and provides separate assessments in listening, writing, reading, grammatical structures, and vocabulary. Other learner corpus research projects also rely on this methodology (e.g. COREFL –CORpus of English as a Foreign Language). For technical reasons, speech is unfortunately not tested. Because testing all competences is very time-consuming and would have discouraged the subjects from participating, only the listening comprehension was tested. It was picked out of other competences for the simple reason that this study focuses on oral perception and production.

Participants first had to complete a vocabulary placement test in which a collection of real and invented verbs was presented. They were asked to decide whether these words existed or not. Immediate feedback was given after completing this placement test. A self-assessment questionnaire about the chosen competence was then administered. The difficulty of the listening comprehension test depended on the estimated proficiency level obtained from the placement test

³⁹The original Dialang Project was carried out with the support of the commission of the European Communities within the framework of the SOCRATES programme, LINGUA 2. It is now funded and maintained by Lancaster University. It can be accessed from: <http://dialangweb.lancaster.ac.uk/>

and on the results of the self-assessment questionnaire. After completing the test, participants received feedback regarding their level in the tested competence and were informed whether their self-assessment matched the test results (Zhang & Thompson, 2004). All participants that obtained a B2 or higher proficiency score in all three tests were enabled to take part in the study. Results obtained to each test and for each selected participant are reported in Table 1 below.

Table 1: Results obtained to all tests by each selected participant. For more visibility, B2 level is indicated in red, C1 level in blue, and C2 level in green

| Subject | Placement test | Self-assessment | Listening comprehension |
|---------|----------------|-----------------|-------------------------|
| S1 | C1 | C1 | B2 |
| S2 | C1 | C1 | B2 |
| S3 | C2 | B2 | C2 |
| S4 | C2 | B2 | C2 |
| S6 | C2 | C2 | C1 |
| S7 | C1 | C2 | C2 |
| S8 | C2 | B2 | C1 |
| S10 | C1 | C2 | B2 |
| S13 | C2 | C1 | C2 |
| S14 | C1 | C1 | B2 |
| S15 | C2 | B2 | B2 |
| S16 | B2 | C1 | B2 |
| S17 | C2 | C2 | C2 |
| S20 | C2 | C2 | C1 |
| S22 | C1 | C1 | B2 |
| S23 | C2 | C1 | C2 |
| S24 | C2 | C2 | B2 |
| S27 | C2 | C2 | B2 |
| S28 | C1 | C2 | B2 |
| S29 | C2 | C2 | B2 |

Because oral expression could not be tested, participants self-reported their speaking proficiency and degree of accentedness in the language background questionnaire (Dmitrieva *et al.*, 2020)⁴⁰. Participants' average self-reported degree of accentedness in English was 'fairly low' (3 on a 7-point scale ranging from 'weak' to 'strong'). Self-reported speaking proficiency was 'high' on average (6 on a 7-point scale ranging from 'very low' to 'very high'). Participants

⁴⁰See Appendix C.2

also reported spending 15.9 hours listening to English on average ([SD] = 12.9), and 9.3 hours speaking English ([SD] = 14.3), per week. 90% of the participants declared listening mostly to American English and 10% to British English. 60% of them identified their accent as being more American and 40% as more British.

4.2 Speech tasks

All tasks were designed and conducted using PsychoPy (v2021.2.3)⁴¹. Each subject was given 15€ for their participation to the whole study.

4.2.1 Reading task

Participants first conducted a reading/control task in which they read the same stimuli as those recorded by the model speakers. This served as the baseline production of each subject. Half of the subjects started the task in French and the other in English. This was meant to avoid any vocal fatigue effect that can be induced when speaking for a certain amount of time, and that could contribute to a higher degree of creakiness. Stimuli appeared on the screen one by one, in a randomized order. Participants were instructed to read the sentence as naturally as possible when it appeared on the screen. They were asked to click on the space bar once they were done reading the sentence to proceed with the next one. This was repeated until they were informed the task had been completed. It took approximately 3 min to complete the whole task. Segmentation for these recordings followed the same pattern as that made for the recordings of the model speakers, except that the *repetition* tier was replaced with a tier named *sentence* which contained the number corresponding to the sentence that was produced (e.g. sentence 14 was labelled as P14, sentence 9 as P9, etc.)

4.2.2 Imitation task

Subjects that started the reading task in French also started the imitation task in French, and vice versa. Each subject listened to the presented stimulus once through headphones. Stimuli were presented in a randomized order in each language and for each subject for the same reasons

⁴¹Code for experiments will be made accessible on: <https://github.com/LeaBurin/PsychoPyExperiments.git>

mentioned above. Participants had been instructed beforehand to repeat and imitate as closely as they could the production of the model speakers. No further instructions as to what linguistic features they were supposed to imitate were provided. They were asked to click on the space bar one once they were done repeating the sentence to proceed with the next one. This was repeated until they were informed the task had been completed. It took approximately 20 mins to complete the whole task. Segmentation for these recordings followed the same pattern as that made for the recordings of the reading task, except for the fact we added a tier named *M_speaker*, and another one named *VQ*⁴². This was made to make sure we knew what model speaker they were repeating, and whether the sentence they were repeated was produced in either non-creaky or creaky voice.

4.2.3 Judgment rating task

A judgment rating task was conducted after the imitation task. We evaluated the attitude towards the production of creaky vs. non-creaky voice across language and gender.

Based on Bayard *et al.* (2001), participants had to rate their impression of the person's voice that was presented to them on 4 six-point semantic scales (pleasant, attractive, powerful, and educated)⁴³ with anchors of 'not at all' and 'very'. In each language, 3 same sentences produced in both qualities by 8 model speakers (2 American male, 2 American female, 2 French male, 2 French female) were used, for a total of 48 sentences to rate. These sentences were selected from the stimuli previously recorded. They were presented in a randomized order for each subject and both languages were mixed up. This was meant to prevent subjects from expecting to hear one specific voice, sentence, or language (considering the low amount of speakers and sentences used in this task). All traits were displayed on the same page on which the stimulus was played. Participants could listen to the stimulus as many times as they wanted by clicking on a red square, as can be observed in Figure 8. They were asked to click on the space bar once they were done rating one voice to proceed with the next one. This was repeated until they were informed

⁴²NC was indicated when the sentence was produced in non-creaky voice, and C when it was produced in creaky voice.

⁴³In their study, Bayard *et al.*, 2001 describe a series of evaluation of gender pairs of different voices speaking different English dialects. They use many more traits but we only used those they defined as 'voice quality traits' considering we were also interested in the evaluation of creaky vs non-creaky voice. We excluded the 'strong' trait because the semantic content of that word was too close from that of 'strong'.

the task had been completed. It took approximately 12 min to complete the whole task.

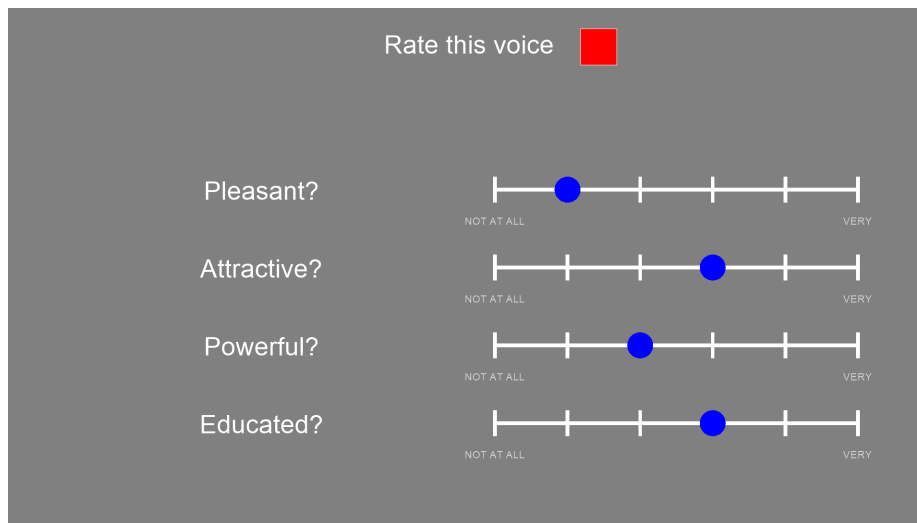


Figure 8: Screenshot of what participants would see during the judgment rating task.

The procedure of the whole experiment, which includes a language proficiency test and three different speech tasks, was described in this section. Based on previous findings, we discuss our different hypotheses in the next section.

5 Hypotheses

We investigated the influence of language and gender on the evaluation and accommodation of creaky voice. Based on previous findings, we detail our hypotheses for accommodation of creaky voice in section 5.1, and those for perception of creaky voice in section 5.2.

5.1 Accommodation

H1: Language - There is an effect for language (French/English) on the success of imitation of creaky voice on the last word of the prosodic unit.

Pillot-Loiseau *et al.* (2019) observed that the amount of creaky voice in read speech would depend on the speaker's native language. Native speakers of English would creak more than

native speakers of French. They would also produce longer creaky occurrences. The proportion of creaky voice would be greater in the French speakers' L2 when they would read in English.

H2: Gender of the model speaker - There is an effect for gender of the model speaker (cisgender male/cisgender female) on the success of imitation of creaky voice on the last word of the prosodic unit.

Women are generally found to converge more than men (Namy *et al.*, 2002; Babel, 2012; Babel *et al.*, 2014), and both men and women appear to accommodate more towards men than towards women (Willemyns *et al.*, 1997; Bilous & Krauss, 1988; Gallois & Callan, 1988; Namy *et al.*, 2002). Contradictory results have, however, also been observed, more particularly in the case of creaky voice: men were found to converge more than women, and more towards women than towards men (Wright *et al.*, 2019).

5.2 Evaluation

H1: Language - There is an effect for language (French/English) on the evaluation of creaky voice.

Creaky voice has become a common feature in the speech of young American women these past years (Lefkowitz & Sicoli, 2007; Yuasa, 2010, Podesva, 2013). Phonation settings in French are often described as being nasalized or breathy, but there is no mention of creaky voice apart during hesitation manifestations (Benoist-Lucy & Pillot-Loiseau, 2013). Experience shows that French learners are also predominantly exposed to American varieties as compared to British ones (e.g. media)⁴⁴.

H2: Gender of the model speaker - There is an effect for gender of the model speaker (cisgender male/cisgender female) on the evaluation of creaky voice.

Brown *et al.* (1974) found that high pitch resulted in speakers being judged as less competent and benevolent. They were judged as less truthful, less emphatic, less “potent” (smaller, thinner, faster), and more nervous by Krauss *et al.* (1979). Males speaking with a lower voice, therefore

⁴⁴ Answers given in the language background questionnaire confirmed this observation. As mentioned in section 4.1, 90% of the participants declared listening mostly to American English and only 10% to British English.

possibly with a creaky voice, are often perceived more positively than men speaking with a higher voice (Esling, 1978; Yuasa, 2010). Conflicting results were observed for female speakers speaking with a creaky voice. Young Northern Californian women speaking with a creaky voice were perceived as being more professional, educated, non-aggressive, and genuine in Yuasa’s (2010). Conversely, negative judgments have also been associated with the use of creaky voice in the speech of American women. Women who would exclusively use modal voicing were judged as being more educated, trustworthy, and hireable as compared to those who would speak with a creaky voice in Anderson *et al.*’s (2014). Lee (2016) found that both male and female speakers were perceived as less educated, less intelligent, less confident, less feminine, more masculine, and more hesitant when using creaky voice phonation as compared to the modal register.

Table 2: Hypotheses and predictions.

| Effect | Language | Gender |
|----------------------|--|---|
| <i>Accommodation</i> | Greater convergence towards American English speakers | Greater convergence towards male speakers |
| <i>Evaluation</i> | More positive evaluation towards American English creaky voice | ? (empirical question) |

Our different hypotheses were presented in this section and are summarised in Table 2. Our predictions are based on the studies cited in this section. In the next section we detail the different measurements we made. Software used to extract the acoustic and EGG measures are described in section.

6 Measurements

Based on Keating *et al.* (2010) and Esposito’s (2012), fundamental frequency, duration, eight spectral measures, and two EGG measures were made on each creaky vowel. Of the eight

spectral measures, one measure, Cepstral Peak Prominence (CPP), is a measure of periodicity. The other seven are spectral measures H1*, H2*, H1*-H2*, H1*-A1*, H1*-A2*, H1*-A3*, H2*-H4*. Esposito (2012) studied phonation in White Hmong and observed correlations between the acoustic and EGG measures she made. Keating *et al.* (2010), who studied phonation categories within and across language, found that several acoustic measures would differentiate them within each language, but that only H1*-H2* would do so in all four studied languages. OQ - obtained through EGG measurements - also appeared to differentiate in most of the languages. Phonation types with the same descriptive name hence appear to differ significantly along several dimensions. They concluded that “these consistent cross-linguistic differences [...] suggest that language/speaker/recording differences in voice quality are larger than phonation category differences” (Keating *et al.*, 2010: 198).

This subsection summarizes the main findings in the field for the acoustic dimensions we investigated.

The most common acoustic measurements used to analyse phonation were described in Chapter I. We report the main findings in the field for acoustic and EGG dimensions we investigated in section 6.1. Although this study focuses on creaky voice, measurements of modal voice and breathy voice may sometimes be reported as a comparison⁴⁵. Cross-linguistic and cross-gender differences may also be introduced for both language and gender are taken as independent variables in this study. Software used to extract both acoustic and EGG measurements are described in section 6.2.

6.1 Acoustic measures

6.1.1 Duration

Creaky vowels have been found to have a greater overall duration than their modal counterparts in Jalapa Mazatec (Silverman *et al.*, 1995).

⁴⁵For more studies mentioning breathy voice, see Fischer-Jorgensen, 1967; Bickley, 1982; Klatt & Klatt, 1990; Hillenbrand *et al.*, 1994; Gordon & Ladefoged, 2001; Gobl & Ni Chasaide, 2010, among others.

6.1.2 Fundamental frequency

f_0 is considered to be the main acoustic parameter differentiating between male and female speakers. A longer vocal tract and longer vocal folds will inevitably give rise to lower resonant frequencies (Fant, 1970). Considering that, during puberty, women's vocal folds length increases by approximately 34%, while they increase almost twice as much in men, male speakers' f_0 decreases much more than that of female speakers (Jenkins, 1998 ; Abitbol *et al.* 1999). The difference in the f_0 range between men and women is, however, only partially due to physiological contrasts. As stated by Simpson (2009: 625), “[the] average fundamental frequency of the voice is in part learned”, meaning that social behaviours can also impact this difference in the f_0 range. According to Ohala (1983), there exists a *Frequency Code* that accounts for cross-language and cross-cultural similarities in pitch use. It states that pitch variations can, for instance, signal different social attitudes (e.g. dominance, submission, assertiveness, politeness).

Cross-language differences have also been observed. Yamazawa & Hollien (1992), and Yuasa (2010), observed that American women exhibited a significantly lower mean speaking f_0 than their Japanese counterparts. Japanese speakers appeared to use greater speaking f_0 variability than their American counterparts (Yuasa, 2010). Ohara (1999) observed that Japanese female speakers would use a higher f_0 mean when speaking Japanese, rather than when speaking in English when addressing a professor a friend. Male Japanese speakers' mean f_0 did not significantly differ across language. Pépiot (2013) observed that mean f_0 was higher for female speakers, as compared to male, in both French and American English. f_0 range was found to be significantly greater for female speakers in French, but not in American English.

Creaky voice has been associated with the lowering of f_0 in many languages (Fischer-Jorgensen, 1967; Blomgren *et al.*, 1998; Gordon & Ladefoged, 2001). Yet we saw that this lowering effect is not systematic and that not all sub-types of creaky voice can have a low f_0 . Interestingly, several studies have demonstrated that there can be no significant f_0 differences between male and female speakers of American English in the production of vocal fry, which means that both groups can reach very low and similar qualities, independently of physiological constraints (Blomgren *et al.*, 1998; Chen *et al.*, 2002).

Although f_0 and phonation types seem to be correlated in many languages, some studies

have demonstrated that, for some speakers, there could be no correlation at all (Holmberg *et al.*, 1995; Epstein, 2002). What is considered as high or low for one speaker may not be considered as such for another speaker. The fairly constant presence of inter- and intra-speaker variability must not be ignored or neglected.

6.1.3 Spectral tilt

One cannot analyse study phonation without analysing spectral tilt measures. $H1^*-H2^*$, one of the most common spectral measure, varies continuously between more constricted voice qualities (manifested by lower $H1^*-H2^*$ and OQ) and less constricted voice qualities (manifested by higher $H1^*-H2^*$ and OQ).

As OQ increases, energy in the first harmonic (and thus $H1^*-H2^*$) is assumed to increase, and this increase is the presumptive cause of the change in voice quality.

Kreiman *et al.* (2012: 2625)

This is in line with the fact that creaky voice has been found to be highly correlated with a relatively strong H2 (as compared to H1) or with low H1-H2 values (Epstein, 2002; Gobl & Ní Chasaide, 2010), resulting in a more positive spectral slope (Stevens, 1977; Gordon & Ladefoged, 2001) and dominance in the higher frequencies (Avelino, 2010). Garellek & Seyfarth (2016) found that creaky vowels also had an overall lower $H2^*-H4^*$, along with $H1^*-H2^*$. However, we saw that there are several types of creaky voice, each having its own set of characteristics. All types of creaky voice seem to have a decreased (flat) spectral tilt, except for non-constricted creak that was observed as having higher H1-H2 values than modal voice (Keating *et al.*, 2015).

Other spectral measures showed that more skewing and abrupt changes in the shape of the glottal pulse are characteristics of some types of creaky voice. This is manifested by low $H1^*-A2^*$ values (Garellek & Keating, 2011), and is due to the airflow building up gradually as the vocal folds open, then dropping suddenly when the vocal folds close abruptly (Epstein, 2002). The skewed shape of the pulse, paired with the abrupt changes in its shape, are correlated with an increase in the amplitude of the higher frequency harmonics in the spectra of the source and the speech output (Bickley 1982; Hanson 1997; Ní Chasaide & Gobl, 1997; Avelino, 2010).

Spectral tilt differences have also been observed between male and female speakers. Typical OQ for female and male speakers are 60% and 50%, respectively, which leads to a difference of about 3 dB for the relative amplitudes of H1 and H2 between the two groups (Hanson & Chuang, 1999). This difference is less than that observed by Klatt & Klatt (1990), or by Henton & Bladon (1985) who, respectively, found a 5.7 dB and 6 dB difference. However, the trend is in agreement with the fact that female speakers tend to have larger relative H1 amplitudes, suggesting they have larger OQ, hence a breathier voice quality. Pépiot (2014) analysed and compared the production of words and pseudo-words by 10 Northeastern American English speakers (5 females, 5 males) and 10 Parisian French speakers (5 females, 5 males). He observed, in both languages, significant gender-based differences in H1-H2 intensity measurements in open vowels. H1-H2 appeared to be significantly greater in female speakers, suggesting a greater OQ, hence a breathier voice quality. Male American English speakers were found to produce significantly lower H1-H2 values than male French speakers, indicating a very low OQ. As indicated by Pépiot (2014: 308), “[S]uch results support the claim that female speakers’ breathy voice quality would have a physiological origin, whereas male speakers’ use of creaky voice would rather be socio-phonetic and language-dependent”. Gender-based differences were also observed for H1*-A3*, with an average of 23.4 dB for female speakers and of 13.8 dB for male speakers of American English. This indicates that female speakers have weaker high-frequency component content in the speech signal. Such a difference in tilt is easily perceived, meaning that spectral tilt may be of importance in differentiating male from female voices. H1*-A1 showed an average difference of 3 dB, indicating that female speakers tend to have a weaker F1 amplitude and that the presence of a posterior glottal chink persisting through the entire glottal cycle is more common for female than for male speakers. Female speakers also displayed more noise than male speakers in the frequency range of F3 (Hanson & Chuang, 1999). Esposito (2006), in her studies on Santa Ana del Valle Zapotec (SADV), found that the three phonation types used in this language (breathy, modal and creaky) were distinguished by H1-H2 for women and H1-A3 for men. These differences were later reinforced by EGG data and suggest that phonation types may be realised differently by men and women (Esposito, 2012).

Large-scale studies have demonstrated that spectral tilt is one of the most reliable measure

discriminating between phonation types in a number of languages (Gordon & Ladefoged, 2001). Esposito (2006) analysed different languages (e.g. SADV, Gujarati, etc.) and found that H1*-A3* and H1*-H2* were both relatively efficient at differentiating between phonation types within languages. However, she also found in her study of White Hmong (2012) that H2*, H1*-A1*, H1*-A3*, and H2*-H4* did not distinguish any of the phonation types (breathy, modal, and creaky). Garellek & Keating (2011) observed that both breathy and laryngealized phonation differed from modal phonation on the four measures tested (H1*-H2*, H1*-A1*, H1*-A2*, CPP) in Jalapa Mazatec. Avelino (2010) compared the spectra of modal and laryngealized vowels in Yalagag Zapotec (YZ) and observed an increase in the amplitude of A1 for both phonation types. Moreover, the spectral tilt observed in H1-A3 showed “the expected greater magnitude for modal phonation than that of laryngealized vowels, as the gradual adduction of the vocal folds would excite frequencies close to f_0 .” (Avelino, 2010: 273). Keating *et al.* (2010) compared contrastive phonation in 4 languages, namely Gujarati (modal vs. breathy), Hmong (modal vs. breathy vs. creaky), Mazatec (modal vs. breathy vs. creaky) and Yi (tense/lax), and observed that only H1*-H2* enabled the distinction between phonation types in all four languages. They also observed that language/speaker differences in voice quality were larger than phonation category differences across language. No significant interactions with speaker gender was found: in each language, male and female speakers would make the phonation contrasts in similar ways.

Spectral tilt measurements appear to reliably distinguish between phonation types. One has, however, to be aware that “raw values of spectral tilt measures do not index a precise voice quality; one person’s creaky voice can have an average H1-H2 of -2 dB while another person’s creaky voice averages 5 dB” (Garellek, 2019: 89). Since voice quality varies in a continuous way, some degree of inter-individual variability will necessarily be observed. Fairly wide variations in spectral tilt among speakers have, indeed, been reported, meaning that spectral measures can be speaker-dependent (Hanson, 1997; Kreiman *et al.*, 2012).

6.1.4 Cepstral Peak Prominence (CPP)

Relative to modal phonation, creaky voice results in an increased amount of noise, which is due to an irregular pitch. If f_0 is not regular, then the signal’s noise increases (Garellek,

2019). The clarity of individual pitch pulses on a waveform appears to be lower for these two phonation types (Gordon & Ladefoged, 2001). CPP values are lower in the production of creaky voice which is related to that aperiodicity (Garellek, 2019).

6.1.5 Summary of the acoustic measurements

Table 3: Summary of the acoustic measurements extracted with VoiceSauce (from: Shue *et al.*, 2011). (*) indicates that the harmonic and spectral amplitudes were corrected for formant frequencies and bandwidths.

| Measurement | Description |
|-----------------------|---|
| CPP | Cepstral Peak Prominence (in dB) |
| H1* | The amplitude of the first harmonic (in dB) |
| H2* | The amplitude of the second harmonic (in dB) |
| H1*-H2* | The amplitude of the first harmonic minus the amplitude of the second harmonic (in dB) |
| H1*-A1* | The amplitude of the first harmonic minus the amplitude of the first formant peak (in dB) |
| H1*A2* | The amplitude of the first harmonic minus the amplitude of the second formant peak (in dB) |
| H1*-A3* | The amplitude of the first harmonic minus the amplitude of the third formant peak (in dB) |
| H2*-H4* | The amplitude of the second harmonic minus the amplitude of the fourth formant peak (in dB) |
| STRAIGHT f_0 | Fundamental frequency (in Hz) calculated using the STRAIGHT algorithm (Kawahara <i>et al.</i> , 1999) |
| Duration | Vowel duration (in ms) calculated by subtracting the seg_End values with the seg_Start values |

A summary of the acoustic measures used in this study is given in Table 3. All values were extracted using VoiceSauce. This software will be described in a following section.

6.2 Electroglottographic measures

6.2.1 Closed quotient

As mentioned in a previous chapter, $CQ = 1 - OQ$ (Henrich, 2001). The reader can link reported findings in the previous section for H1*-H2*/OQ-CQ. A large OQ/small CQ will indicate a breathier quality, while a small OQ/large CQ will indicate a creakier quality. EGG analyses conducted on different languages by Keating *et al.* (2012) confirmed that creaky voice is produced with large CQ values. CQ was also found to be significantly greater in creaky phonation as compared to that in breathy phonation in White Hmong (Esposito 2012).

6.2.2 PIC (Peak Increase in Contact)

In her study of White Hmong, Esposito (2012) observed high PIC values for breathy phonation and low PIC values for creaky phonation⁴⁶. The same observation was made by Keating *et al.* (2012) in their acoustic and EGG comparison of various languages.

7 Software

7.1 EggWorks

| Options | | |
|--------------------------------------|--------------------------------------|---|
| EGG Channel: | <input type="text" value="0"/> | File's EGG channel, zero based. Ignored for single channel files. |
| Smooth Window (points): | <input type="text" value="0"/> | EGG data smoothing number of points |
| Velocity Smooth Window (ms): | <input type="text" value="0.5"/> | EGG velocity data smoothing window |
| Threshold percent (%): | <input type="text" value="25"/> | Threshold percentage for the basic analysis |
| DC Shift window (ms): | <input type="text" value="8"/> | Used only for the Henry Tehrani method (for estimating the DC movement). |
| Invert EGG channel: | <input type="button" value="No"/> ▾ | |
| VoiceSauce Compatible output: | <input type="button" value="Yes"/> ▾ | VoiceSauce requires a specific file format generated, with 1 msec interval redundant values for each frame, with output extension of ".egg" |

Figure 9: EggWorks parameters as found on the software page.

Electroglottographic signals were extracted and automatically analysed using EggWorks (Tehrani, 2009). The first step in processing was the inversion of the recordings made at Univer-

⁴⁶She refers to this value as DECPA and not PIC in her study, but these labels account for the same measure, as we will described in the following section.

sité Paris Cité (recordings were made via a laptop soundcard and appeared to be inverted). The program then calculates different measures for each glottal pulse it can find, throughout each entire file. To match VoiceSauce (described in the following sub-section), EggWorks interpolates values to 1-msec intervals and outputs either a text file or an .egg file for each EGG signal file. You can then ask VoiceSauce to include the EggWorks outputs (the .egg files) as additional parameters for them to be simultaneously analysed with the acoustic parameters selected. Default values were used for all other settings⁴⁷. The different options are showed in Figure 9. Output measurements are represented in Table 4. Measurements marked with an asterisk, namely CQ_H and peak_Vel, are those used in this study. Peak_Vel is often referred to as PIC (Keating *et al.*, 2012). It corresponds to the DECPA (Derivative-EGG Closure Peak Amplitude) measure used by Michaud (Michaud, 2004). We will only refer to this measure as PIC in the following.

All closed quotient measurements made by EggWorks are illustrated in Figure 10. Following Esposito's (2012), the CQ by Hybrid method (CQ_H) was selected for this study because the negative peak in the EGG derivative (DEGG) appeared not always well-defined. The method is said to be hybrid for the edges of the contacting phase of the glottal cycle are defined using two different methods (Orlikoff, 1991; Howard, 1995). The threshold was fixed at 25% (Orlikoff, 1991).

[with this method] the beginning of the contact phase is defined as the positive peak in the first derivative of the EGG signal, while the end of the contact phase is based on a fixed threshold, in this case 25% of the difference between the minimum and maximum amplitude values in each cycle of the EGG signal.

Esposito (2012: 470)

The black curve corresponds to the EGG signal and the blue curve to the DEGG signal. While CQ represents the full cycle of the EGG, CQ_H is defined by the positive peak of the DEGG signal on the left, and by the negative peak of the DEGG signal determined in respect to the 25% threshold showed in green. At the left boundary of CQ_H, CQ takes the value of 1, while it reaches 0 at the right boundary.

⁴⁷Marc Garellek, personal communication.

Table 4: EggWorks measurements output as taken directly from the software webpage. Measurements marked with an asterisk are those used in the study.

| Measurement | Description |
|----------------------|---|
| FrameStart | cycle start time |
| FrameStop | cycle stop time |
| TCstart | closure start time |
| TCend | closure end time |
| peak_Vel* | cycle peak velocity value |
| peak_Vel_Time | cycle peak velocity time |
| min_Vel | cycle minimum velocity value |
| min_Vel_Time | cycle minimum velocity time |
| SQ1 | SQ1 is the time of 10% above the minimum value of each cycle (closing slope) |
| SQ2 | SQ2 is the time of 90% above the minimum value of each cycle (closing slope) |
| SQ3 | SQ3 is the time of 90% above the minimum value of each cycle (opening slope) |
| SQ4 | SQ4 is the time of 10% above the minimum value of each cycle (opening slope) |
| ratio | the ratio of $(SQ2-SQ1)/(SQ4-SQ3)$ |
| CQ | standard percentage method, using a preassigned percent of the cycle height for TCstart |
| CQ_H* | hybrid method, using the peak velocity time for TCstart, going across until crossing the next cycle |
| CQ_PM | peak velocity to min velocity time, using the cycle's velocity trace peak and min times as the guide |
| CQ_HT | Henry Tehrani method, incorporating the DC component of the EGG signal into the calculation. The peak velocity is TCstart time, then following the DC contour of the EGG signal to cross the down sloping sets the TCend. |

PIC can be observed in Figure 11. The black curve represents the EGG signal and the blue curve the DEGG signal. As defined by Michaud (2004: 1), “[it] is a measurement at one single point in time for each period: it corresponds to the highest speed in increase of vocal fold contact surface which is reached at the glottis-closure-instant”. PIC and PDC correspond to number 2 and 5 in Figure 6 (section 1.3), respectively.

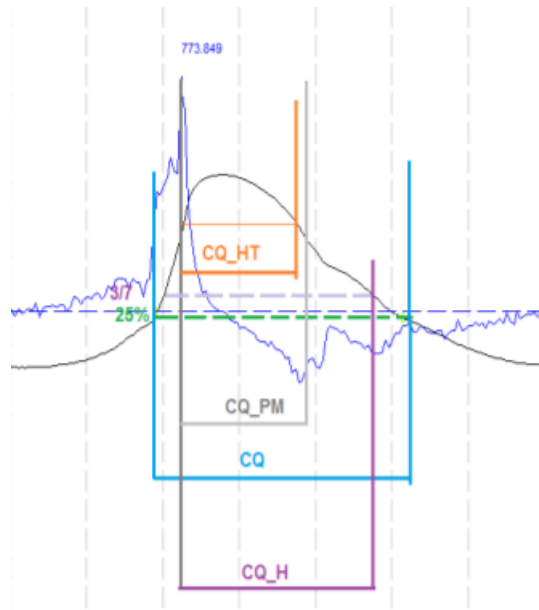


Figure 10: Measures output from EggWorks (Tehrani, 2009) showing CQ, CQ_H, CQ_PM and CQ_HT. This figure was extracted from Keating *et al.*, 2012.

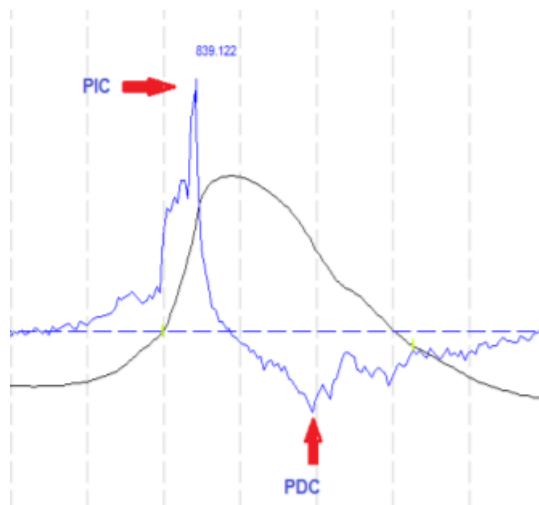


Figure 11: The derivative of the signal on which is indicated the peak decrease in contact (PDC) and the peak increase in contact (PIC), which corresponds to the peak_Vel measure in EggWorks. This figure was extracted from Keating *et al.*, 2012.

7.2 VoiceSauce

All measurements were extracted using VoiceSauce. VoiceSauce is a Matlab-implemented application which provides automated voice measurements over time from audio recordings (Shue *et al.*, 2011). For each input, you can control whether all the data is output, or whether you want

it averaged into chunks. If you select no subsegments in the output-to-text settings, then it will write all the data, with one measurement per interval specified by the Frame Shift parameter that you can see in Figure 12 below. We can see that the default Frame Shift value has not been modified and has been left to 1 ms. For a segment that is 100 ms long, then the output text file will list each measurement on a separate row, resulting in 100 rows for one segment. If the Frame Shift was to be modified to be 10 ms, then it would have 10 rows of the corresponding data. If you decide on using sub-segments then the data will be averaged into n number of chunks. Changing the number of sub-segments to 12 will give you 12 measurements written in separate columns for each segment. Measurements will still be taken at 1 ms but averaged across 12 time points of the segment's duration. The mean measurement averaged over the whole course of the segment's duration will also be written⁴⁸. Non-modal phonation does not necessarily occur over the entire course of vowels. Esposito (2004), for instance, found that creaky vowels only contrast from modal ones at the ends of vowels in Zapotec. Creaky voice might occur at different locations in both French and American English vowels. For this reason, measurements were made over the entire vowel as well as averaged over each third of the vowel. VoiceSauce only outputs the time at which the segment starts and ends. Duration was calculated for each segment by subtracting the Seg_End value to the Seg_Start value.

VoiceSauce parameters are displayed in Figure 12. Parameters framed in red were modified. The STRAIGHT method (*Speech Transformation and Representation based on Interpolation of weiGHTed spectrogram*) was used to analyse f_0 for it is fairly robust in the face of f_0 irregularity, therefore of non-modal phonation, and is widely used in the field (Garellek & Keating, 2011; Kreiman *et al.*, 2012; Keating & Garellek, 2015). The default Min and Max f_0 values are used to restrict the f_0 tracking algorithm so that it only searches in a practical range. We decided to lower the Min f_0 value to 30 Hz (the Min f_0 default value is 40 Hz) after extracting the Min and Max values for each vowel and realising that some values were below 40 Hz. Harmonic amplitudes were calculated pitch synchronously using the default values. Default values were used for all other parameters⁴⁹. We selected the tier corresponding to the one for which segmentation was

⁴⁸The complete documentation on VoiceSauce can be found on:
<http://www.phonetics.ucla.edu/voicesauce/documentation/index.html>

⁴⁹Yen Shue, personal communication. For further discussion on VoiceSauce, see Shue *et al.* (2011).

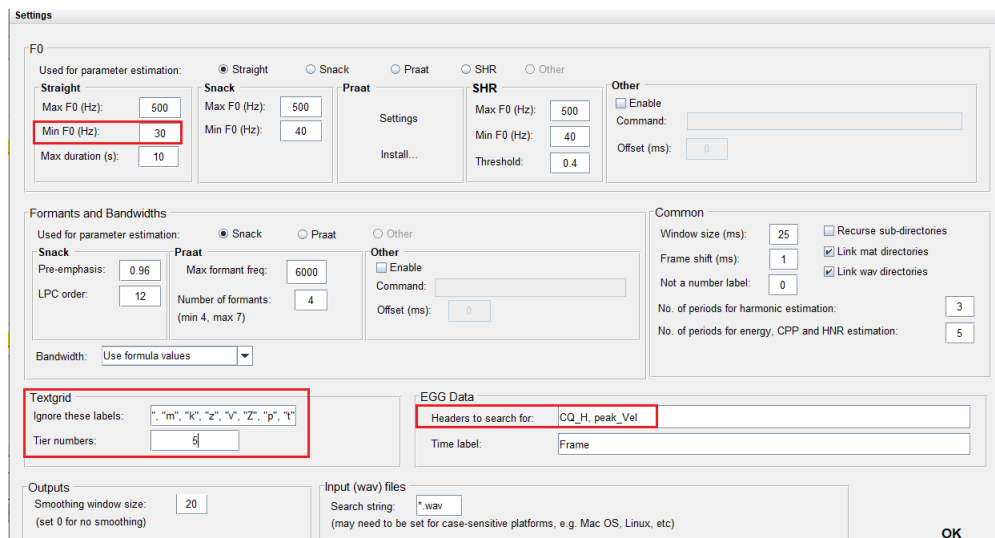


Figure 12: VoiceSauce settings interface. Parameters framed in red were modified. Default values were used for all other parameters.

made at the phoneme level and ignored labels not corresponding to vowels.

8 Conclusion

In the present chapter we provided a description of our whole experimental design. We detailed how recordings were made and how are stimuli were designed and elicited. We discussed participants' selection and the language proficiency test they had to pass in order to take part in the experiment. We described the three different tasks that constitute our experiment and our different hypotheses. We expect an effect of language and gender on the accommodation and evaluation of creaky voice. All measures made were listed in the last section of this chapter. We detailed the different difficulties and obstacles we have encountered, as well as the choices and concessions we made. Results are reported in the following chapters. In Chapter IV we provide a descriptive analysis of creaky voice across language and gender. Convergence analysis will be the topic of Chapter V. Jupyter Notebook 6.4.5 (Anaconda navigator 2.1.1) and RStudio 2022.7.1 were used to conduct the analyses.

CHAPTER IV: A comparative analysis of creaky voice across language and gender

Although the acoustics and physiology of American English creaky voice has been extensively studied and described, little is known about that of French creaky voice. To observe how French creaky voice might differ from American English creaky voice, we analysed all acoustic and EGG measurements extracted from our model speakers' dataset. We used linear mixed-effects models to compare the production of creaky voice across language and gender. Linear mixed-effects modelling offers the advantage to study a combination of independent variables simultaneously and random effects, which can be useful to handle individual differences. The disadvantage is that only one speech feature can be analysed at the same time and multiple similar models will be fit with only the dependent variable changing. Although the number of tokens for Vowel and Vowel type was uneven, we decided to include them in our models as random effects. Even with a low number of tokens a possible relationship between vowel quality and creaky voice could be observed. Iseli *et al.* (2007), for instance, found that $H1^*-H2^*$ values were the lowest for high vowels, suggesting that high vowels are realised with a lower OQ. When possible, we attempted to translate our observations into physiological terms. Principal component analyses (PCA) were performed to observe whether the same set of variables was responsible for the variance observed across language. This method also provides information regarding existing correlations between acoustic and EGG measures.

CHAPTER CONTENT

This chapter provides a descriptive and comparative analysis of French and American English creaky voice. Cross-gender differences are analysed within-language. In the first two sections we describe the pre-processing of our data and data visualization, respectively. In the third section we discuss statistical model selection. We list the different variables we used and how our models were fitted. The acoustics of creaky voice across language and gender is analysed in the fourth section. The analysis of EGG measures (CQ and PIC) is conducted in the fifth section. Finally, principal component analyses are reported in the sixth and last section of this chapter. We analyse the variance observed in our dataset and correlations between acoustic and EGG measures. Cross-gender differences are also analysed within-language.

1 Data processing

All acoustic and electroglottographic measures made in this study were described in the previous chapter. We made those measurements over the entire vowel as well as averaged over each third of the vowel. However, some CQ and PIC values were output as 0, mostly for measurements taken at point 1 or 3. We tested whether we would obtain less 0 values if we made measurements over more or less segments but we did not. We had to decide between excluding the corresponding stimuli or not analysing measurements taken at different points and only focus on measurements made over the entire vowel. We opted for the second option to keep an even number of stimuli per speaker⁵⁰. Among the EGG measurements made over the entire vowel, six occurrences had a mean value equal to 0. All these impossible 0 values were removed from the dataset (Lee *et al.*, 2019).

Once our data had been processed we started observing our data depending on variables of interest. Data visualization is detailed in the next section.

2 Data visualization

We opted for violin plots for data visualization. Violin plots are a combination of boxplots and kernel density plots. They allow the comparison of the distribution of the data across different variables of interest, here Language and Gender. A narrower density represents a lower probability for members of a population to take on a given value, and a wider density represents a higher probability. In other words, there is more chance for a value to occur more frequently on wider regions and less chance for a value to occur more frequently on skinnier regions. The peaks display where values are concentrated. For all violin plots representing independent variables, American English speakers' production is represented on the left side and French speakers' production on the right side. Female speakers' production is represented in blue and male speakers' production in yellow. The dashed line represents the median and the dotted lines the 1st and 3rd quartiles. Other violin plots are used to represent the effect of

⁵⁰According to Yen Shue (personal communication), there might have been something in EggWorks that caused the signal to stop processing. We tried boosting the signal for it to be more prominent and managed to get EggWorks to process up the signal, but only for a few stimuli.

either Speaker or Vowel on the dependent variable. Line plots are used to represent the effect of either Language, Gender, the interaction between both, or Vowel type on the dependent variable (because these are 2-factor variables). For all tables, FM refers to French males, FF to French females, AM to American males, and AF to American females.

Some tendencies observed graphically may not result in statistically significant results. In the next section we discuss statistical model selection.

3 Model selection

Separate mixed-effects models were built to assess cross-language and cross-gender differences on all acoustic and EGG measures. This was done using the function `lmer()` of the *lmerTest* package in 2022.7.1 (Kuznetsova *et al.*, 2017). For all 12 variables, models employed the maximal random effects structure by including intercepts for all random factors⁵¹:

```
lmer(variable ~ Language + Sex + Language:Sex + (1 | Speaker) + (1 |  
Vowel_type) + (1 | Vowel)
```

The different elements the model was fit with are described below.

- **Dependent variable:** acoustic or EGG variable (e.g. f_0 , H1*, H1*-H2*, etc.)
- **Fixed effects:** Language (2 levels: *French and English*), Gender (2 levels: *male and female*), and the interaction between the two (Language*Gender)
- **Random effects:** Speaker (12 levels: *AF_01, AF_04, AF_06, AM_01, AM_02, AM_03, FF_01, FF_02, FF_08, FM_02, FM_05, FM_06*), Vowel (10 levels: *KIT, GOOSE, LOT, TRAP, IL, RUE, ROUE, PÂTE, PLAT*), and Vowel type (2 levels: *Low/High*)⁵²

Some of the models resulted in singular fits, meaning that one variance or more linear combinations of effects were close to zero (Winter, 2020). There was no systematic effect coming from

⁵¹We did not have enough data to include random slopes.

⁵²We did not include Sentence due to the lack of occurrences of each sentence across the dataset. It is also beyond the scope of this dissertation to analyse this effect.

the variable in question so the model was simplified by removing all random effects for which variance was close to 0⁵³. As a result, not all models have the same random effects structure. For each variable, the best model was identified using the `step()` function which includes only significant factors⁵⁴. We will report the best model found along with both fixed and random effects included (indicated in Tables). No fixed or random factors were systematically observed as having a significant effect on the variable under consideration. This might be due to the lack of tokens present in our dataset.

We detailed how we built our linear mixed-effects models to analyse and compare French and American English creaky voice. In the next section we analyse every acoustic measures independently.

4 Analysis of acoustic measures

In this section we provide an analysis of acoustic measures. Analyses of duration, f_0 , H1*, H2*, H1*-H2*, H1*-A1*, H1*-A2*, H1*-A3*, H2*-H4*, and CPP are reported from section 4.1 to 4.10, respectively.

4.1 Duration

The distribution of creaky vowels' duration values across language and gender is displayed in Figure 13. Overall, the range of values is quite similar across language. The distribution of values between male and female American English speakers does not vary much, with a peak around 200 ms for both groups. The distribution is only slightly wider in the speech of male American English speakers as compared to that of their female counterparts. There appears to be more variability in the distribution of duration values when creaky vowels are produced by female French speakers than when they are produced by male French speakers. The peak is located around 190 ms for female French speakers and around 140 ms for male French speakers.

⁵³We used the `ranef()` function from the `lme4` package to identify these random effects.

⁵⁴We used the backward stepwise selection in which the model is first fitted with all variables under consideration and then starts removing the least significant variables one after the other until it only includes significant factors.

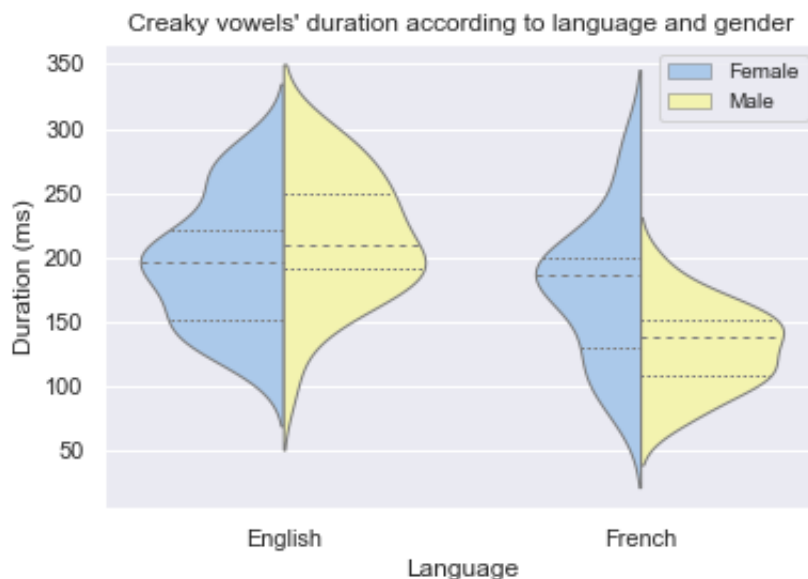


Figure 13: Violin plot representing the distribution of creaky vowels' duration across language and gender.

Table 5: Mean creaky vowels' duration (in ms) across language and gender. Standard deviation is indicated in parenthesis.

| Mean duration and SD (ms) | | |
|---------------------------|----------------|----------------|
| Language | <i>French</i> | 154.01 (50.05) |
| | <i>English</i> | 206.38 (48.92) |
| Gender | <i>Male</i> | 173.48 (59.46) |
| | <i>Female</i> | 186.46 (51.98) |
| Language:Gender | <i>FM</i> | 130.65 (32.0) |
| | <i>FF</i> | 175.81 (54.34) |
| | <i>AM</i> | 216.32 (48.7) |
| | <i>AF</i> | 197.11 (48.06) |

The tense/lax distinction does not exist in French, meaning that vowels are relatively shorter in this language. The last word of the prosodic unit in French is, however, often subject to lengthening (Vaissière, 2002). Another feature that could account for this between-group variability in duration is speech rate. Since vowel duration is correlated with speech rate, it could mean that some speakers read faster than others⁵⁵. Mean creaky vowels' duration values reported in

⁵⁵We used De Jong & Wempe's (2009) script to test speech rate on the full recording for each speaker but it failed to compute the same number of syllable nuclei for each speaker, although they all read the same sentences 5 times.

Table 5 show that, overall, creaky vowels are shorter in French and when produced by male speakers than in American English or when produced by female speakers. Creaky vowels also seem to be shorter when produced by male French speakers as compared to when produced by their female counterparts, while the reverse pattern is observed for American English speakers. The distinction that exists between tense and lax vowels in American English could explain the fact that French creaky vowels are shorter than American English creaky vowels. Tense vowels have been found to be longer than lax vowels⁵⁶ (Laver, 1994).

Table 6: Significant and non-significant effects for duration (Marginal $R^2 = 0.423$ /Conditional $R^2 = 0.707$).

| Backward reduced random-effect table: | | | |
|--|-------------------|----------------|----------------------|
| | <i>Eliminated</i> | <i>AIC</i> | <i>Pr(>Chisq)</i> |
| (1 Vowel_type) | 0 | -186.97 | 0.01760 * |
| (1 Speaker) | 0 | -177.24 | 8.842e-05 *** |
| (1 Vowel) | 0 | -189.13 | 0.06196 |
| Backward reduced fixed-effect table: | | | |
| | <i>Eliminated</i> | <i>F value</i> | <i>Pr(>F)</i> |
| Language:Gender | 0 | 5.7156 | 0.03577 * |

We ran a linear mixed-effects model with logarithmic converted values of raw duration values as dependent variable. Results, reported in Table 6, show that the interaction between Language and Gender, Vowel type, and Speaker, had a significant effect on creaky vowels' duration. The interaction between Language and Gender is illustrated in Figure 14.

The model that was found for $\log(\text{dur})$ was the following:

$$\text{dur}(\log) \sim \text{Language} + \text{Gender} + \text{Language:Gender} + (1 \mid \text{Speaker}) + (1 \mid \text{Vowel_type}) + (1 \mid \text{Vowel})$$

⁵⁶This is also the case for vowels in stressed syllables (Laver, 1994).

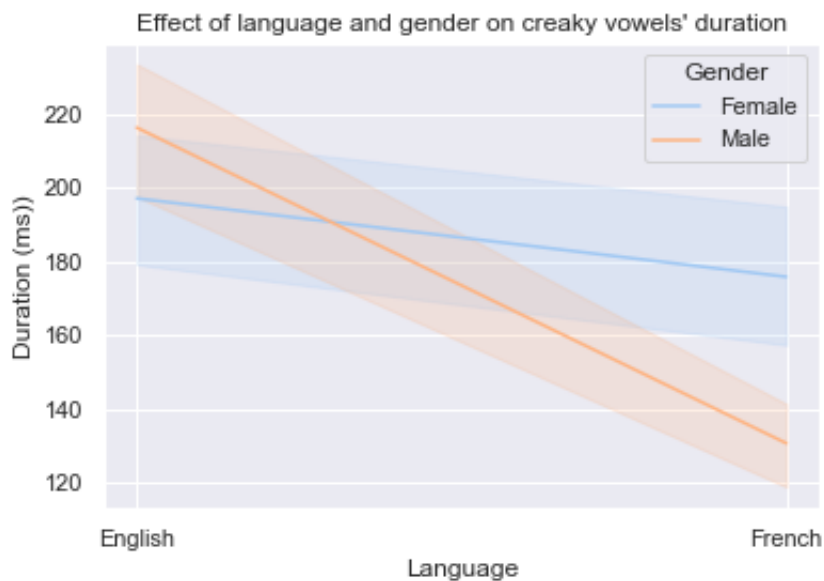


Figure 14: Effect of language and gender on creaky vowels' duration. The shaded areas represent the standard deviation for each group.

Post-hoc comparisons showed that male French speakers produced significantly shorter vowels than both female American English speakers ($p < .01$) and male American English speakers ($p < .01$). There was no significant cross-gender differences observed within-language. Variability in speech rate, or the fact that not all models produced the same sentences, hence not the same vowels, might account for Speaker effect. Figure 15 shows the distribution of creaky vowels' duration across model speakers. We can see that, overall, American English speakers, on the left, seem to produce longer creaky vowels than their French counterparts, and that the distribution is more varied, especially in the production of male American English speakers. This is likely to be due to the tense/lax distinction we mentioned above. There is not a lot of variability in the production of French speakers, apart from FF_02 who stands out. This model probably read slower than other recorded models.

Post-hoc comparisons showed that FM_05 ($M = 115.22 \pm 31.76$) produced significantly shorter vowels than AF_04 ($M = 170.64 \pm 31.04$; $p < .05$), and AF_06 ($M = 191.28 \pm 59.96$), AM_02 ($M = 203.34 \pm 44.67$), AM_03 ($M = 215.13 \pm 48.92$), FF_02 ($M = 227.05 \pm 40.19$), AF_01 ($M = 229.39 \pm 30.18$), and AM_01 ($M = 234.03 \pm 53.91$; $p < .001$); FM_06 ($M = 137.11 \pm 26.74$) produced significantly shorter vowels than AM_02 ($M = 203.34 \pm 44.67$; p

<.05), AM_03 ($M = 215.13 \pm 48.92$; $p < .01$), FF_02 ($M = 227.05 \pm 40.19$), AF_01 ($M = 229.39 \pm 30.18$), and AM_01 ($M = 234.03 \pm 53.91$; $p < .001$); FF_01 ($M = 140.23 \pm 40.55$) produced significantly shorter vowels than AM_02 ($M = 203.34 \pm 44.67$; $p < .05$), AM_03 ($M = 215.13 \pm 48.92$; $p < .01$), FF_02 ($M = 227.05 \pm 40.19$), AF_01 ($M = 229.39 \pm 30.18$), and AM_01 ($M = 234.03 \pm 53.91$; $p < .001$); FM_02 ($M = 141.86 \pm 34.49$) produced significantly shorter vowels than AM_03 ($M = 215.13 \pm 48.92$; $p < .05$), FF_02 ($M = 227.05 \pm 40.19$), AF_01 ($M = 229.39 \pm 30.18$), and AM_01 ($M = 234.03 \pm 53.91$; $p < .01$); and FF_08 ($M = 160.15 \pm 40.71$) produced significantly shorter vowels than FF_02 ($M = 227.05 \pm 40.19$), AF_01 ($M = 229.39 \pm 30.18$), and AM_01 ($M = 234.03 \pm 53.91$; $p < .05$).

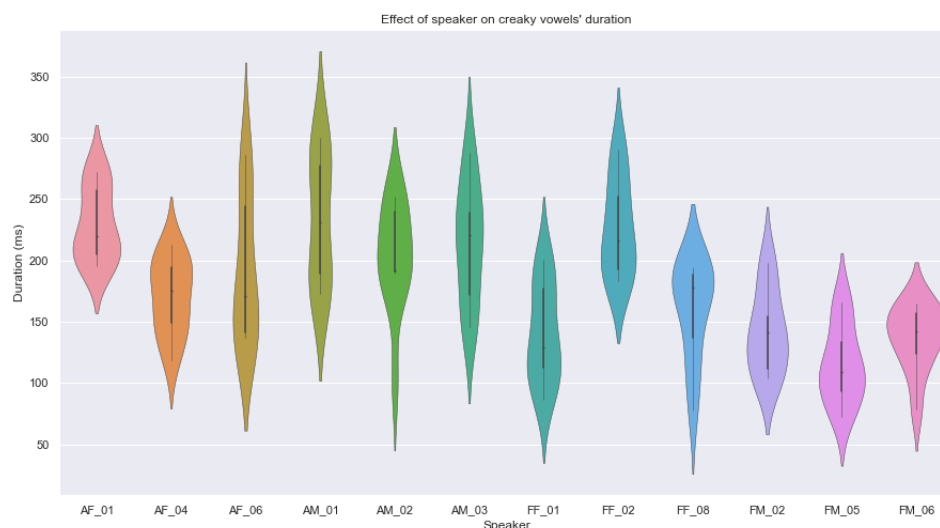


Figure 15: Violin plot representing the distribution of creaky vowels' duration across model speakers.

There is an uneven number of both lax and tense vowels (only valid for American English speakers; $N_{Tense} = 32$; $N_{Lax} = 26$), as well as high and low vowels ($N_{High} = 71$; $N_{Low} = 45$), in the dataset. Some American English speakers might have produced more tense vowels than others, and more in either the High category or the Low category ($N_{Low/Tense} = 2$; $N_{High/Tense} = 30$), which consequently impacted the effect of Vowel type on duration. Phonological context, and more particularly the following segment, might have also played a role for vowels are longer before voiced consonants (Ladefoged, 2006). As observed in Figure 16, we can conclude that, in

our dataset, high creaky vowels' are shorter than low creaky vowels'. This is, however, a dubious observation and we cannot generalize that high creaky vowels are always shorter than low creaky vowels. Moreover, *post-hoc* pairwise comparisons on Vowel type only did not result in significant difference between low and high vowels. Creaky vowels' duration might only distinguish between vowel types when this variable is included in a model along other variables that have a significant effect, but not on its own.

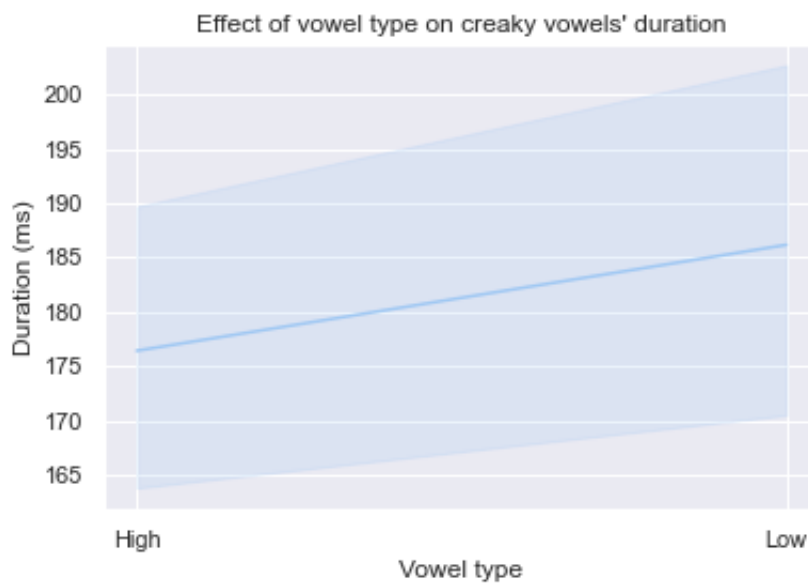


Figure 16: Effect of vowel type on creaky vowels' duration. The shaded area represents the standard deviation.

4.2 Fundamental frequency

The distribution of creaky vowels' f_0 values across language and gender is displayed in Figure 17. Overall, the range of f_0 values is quite similar across language. There is much more variability in the production of female speakers of American English than in the production of their male counterparts. No distinct peak is observed in the distribution of values when creaky vowels are produced by female American English speakers while the peak is located around 65 Hz when produced by male American English speakers. These latter appear to produce creaky vowels with a lower f_0 , but both groups seem to be able to produce these vowels with a similar low

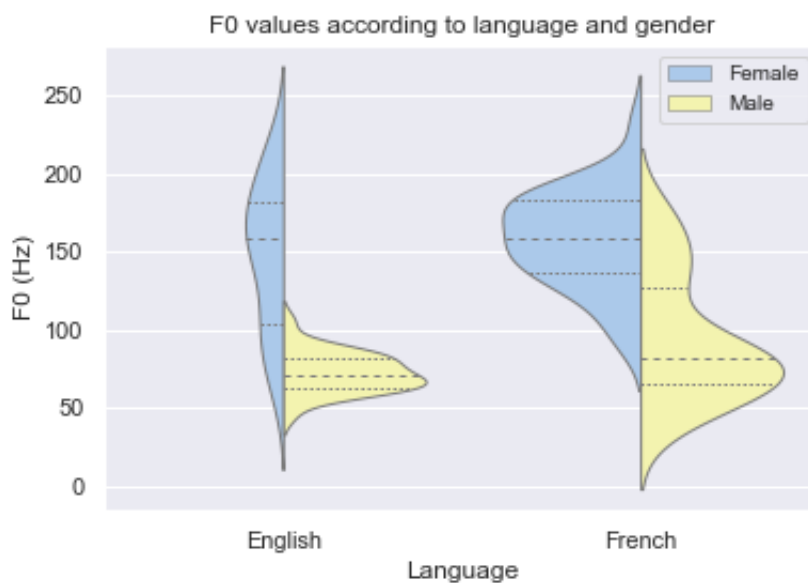


Figure 17: Violin plot representing the distribution of creaky vowels' f_0 values across language and gender.

f_0 . This is not the case for French speakers for which there seems to be a similar amount of variability in both groups, with male speakers producing creaky vowels with an overall lower f_0 , with a peak located around 75 Hz, than female speakers, for which the peak is located around 170 Hz. French female speakers also do not appear being able to reach as low values as French male speakers.

Table 7: Mean f_0 values (in Hz) across language and gender. Standard deviation is indicated in parenthesis.

| Mean f_0 and SD (Hz) | | |
|------------------------|----------------|----------------|
| Language | <i>French</i> | 127.15 (47.69) |
| | <i>English</i> | 111.66 (51.13) |
| Gender | <i>Male</i> | 83.4 (30.79) |
| | <i>Female</i> | 153.02 (39.72) |
| Language:Gender | <i>FM</i> | 93.81 (39.34) |
| | <i>FF</i> | 158.28 (30.88) |
| | <i>AM</i> | 72.98 (12.58) |
| | <i>AF</i> | 147.76 (46.88) |

We can observe in Table 7 that creaky vowels produced by American English speakers have

overall lower mean f_0 than when produced by French speakers. Predictably enough, the same observation can be made for male speakers as compared to female speakers. Creaky vowels also have an overall lower f_0 when produced by male French speakers as compared to when produced by their female counterparts. The same pattern is observed for creaky vowels produced by American English speakers. It is not surprising to observe cross-gender differences for men have longer vocal folds than women, which might give rise creakier phonation.

Table 8: Significant and non-significant effects for f_0 (Marginal $R^2 = 0.52$ /Conditional $R^2 = 0.539$).

| Backward reduced random-effect table: | | | |
|--|-------------------|----------------|----------------------|
| | <i>Eliminated</i> | <i>AIC</i> | <i>Pr(>Chisq)</i> |
| (1 Speaker) | 1 | 1160.2 | 0.4029 |
| Backward reduced fixed-effect table: | | | |
| | <i>Eliminated</i> | <i>F value</i> | <i>Pr(>F)</i> |
| Language:Gender | 1 | 0.6274 | 0.42999 |
| Language | 0 | 5.6902 | 0.01872 * |
| Gender | 0 | 114.7756 | < 2e-16 *** |

The model that was found for f_0 was the following:

$$F0 \sim \text{Language} + \text{Gender}$$

We ran a linear mixed-effects model with f_0 as dependent variable. Vowel and Vowel type were excluded because they did not account for enough variance. Results reported in Table 8 show that the interaction between Language and Gender, as well as the random effect Speaker, were eliminated. However, both Language and Gender had a statistically significant effect on creaky vowels' f_0 . The effects can be observed in Figure 18. Conclusions that we can draw from these observations are that creaky vowels have a significantly lower f_0 in American English than in French, and a significantly lower f_0 when produced by male speakers than when produced by female speakers. In terms of physiology, this could mean that there is more decrease in the activity of the CT muscle and in airflow for American English and male speakers, resulting in increased mass and decreased stiffness of the vocal folds.

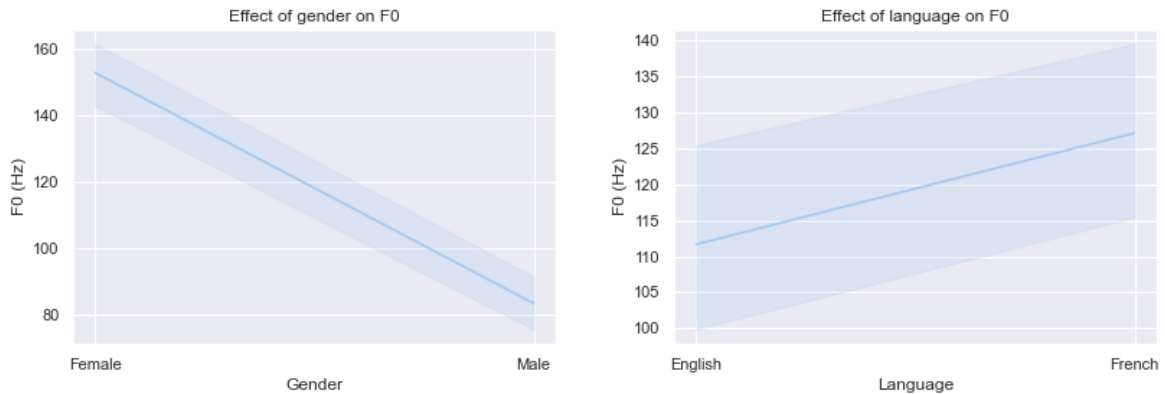


Figure 18: Effect of gender, on the left, and language, on the right, on creaky vowels's f_0 .

The fact that there was no significant gender-based differences observed within-language, and that both male and female American English speakers seem to be able to produce creaky vowels with a similar low f_0 , corroborate observations made by Blomgren *et al.* (1998) and Chen *et al.* (2002). They did not find any significant f_0 differences between male and female American English speakers in the production of vocal fry, and both male and female speakers were found to reach very low and similar qualities.

4.3 H1*

The distribution of creaky vowels' H1* values across language and gender is displayed in Figure 19. Overall, the range of H1* values seems to be wider in French, reaching lower values. The distribution of H1* values between male and female American English speakers is relatively similar, with male speakers producing creaky vowels with a higher H1*, with a peak around 12 dB, and female speakers with a lower H1*, with a peak around 1 dB. The reverse is observed in French in which female speakers produce creaky vowels with a higher H1*, with a peak around 15 dB, and male speakers with a lower H1*, with a peak around 5 dB.

Mean H1* values reported in Table 9 show that mean H1* is higher for creaky vowels produced by French speakers as compared to when produced by American English speakers. It does not

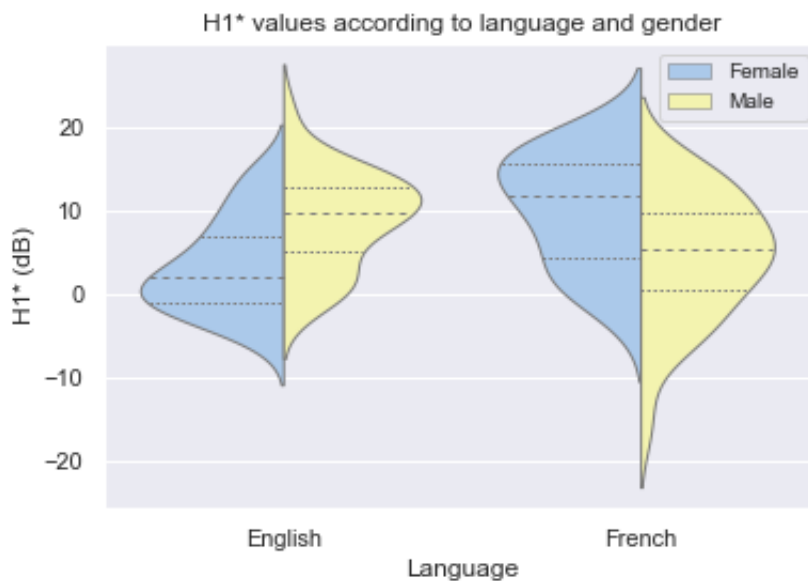


Figure 19: Violin plot representing the distribution of creaky vowels' H1* values across language and gender.

Table 9: Mean H1* values (in dB) across language and gender. Standard deviation is indicated in parenthesis.

| Mean H1* and SD (dB) | | |
|------------------------|----------------|--------------|
| Language | <i>French</i> | 7.38 (7.62) |
| | <i>English</i> | 6.05 (6.24) |
| Gender | <i>Male</i> | 6.68 (6.99) |
| | <i>Female</i> | 6.75 (7.02) |
| Language:Gender | <i>FM</i> | 4.46 (7.51) |
| | <i>FF</i> | 10.12 (6.77) |
| | <i>AM</i> | 8.91 (5.72) |
| | <i>AF</i> | 3.39 (5.56) |

differ as much across gender, although the mean is slightly higher for female speakers. H1* is also higher for creaky vowels produced by female French speakers as compared to when produced by male French speakers, and higher for American male speakers than when produced by their female counterparts.

The model that was found for H1* was the following:

$$H1 \sim (1 \mid \text{Speaker}) + (1 \mid \text{Vowel})$$

Table 10: Significant and non-significant effects for H1* (Marginal $R^2 = 0.186$ /Conditional $R^2 = 0.795$).

| Backward reduced random-effect table: | | | |
|--|-------------------|----------------|----------------------|
| | <i>Eliminated</i> | <i>AIC</i> | <i>Pr(>Chisq)</i> |
| (1 Vowel_type) | 1 | 664.94 | 0.6558 |
| (1 Speaker) | 0 | 763.14 | < 2.2e-16 *** |
| (1 Vowel) | 0 | 688.65 | 3.952e-07 *** |
| Backward reduced fixed-effect table: | | | |
| | <i>Eliminated</i> | <i>F value</i> | <i>Pr(>F)</i> |
| Language:Gender | 1 | 4.3295 | 0.05962 |
| Gender | 2 | 0.0022 | 0.96299 |
| Language | 3 | 0.0057 | 0.94051 |

We ran a linear mixed-effects model with H1* as dependent variable. Results reported in Table 10 show that H1* did not distinguish creaky vowels between neither Language, nor Gender. The interaction between both variables was also found to be not significant. Among the three random effects, Speaker and Vowel had an effect on creaky vowels' H1*.

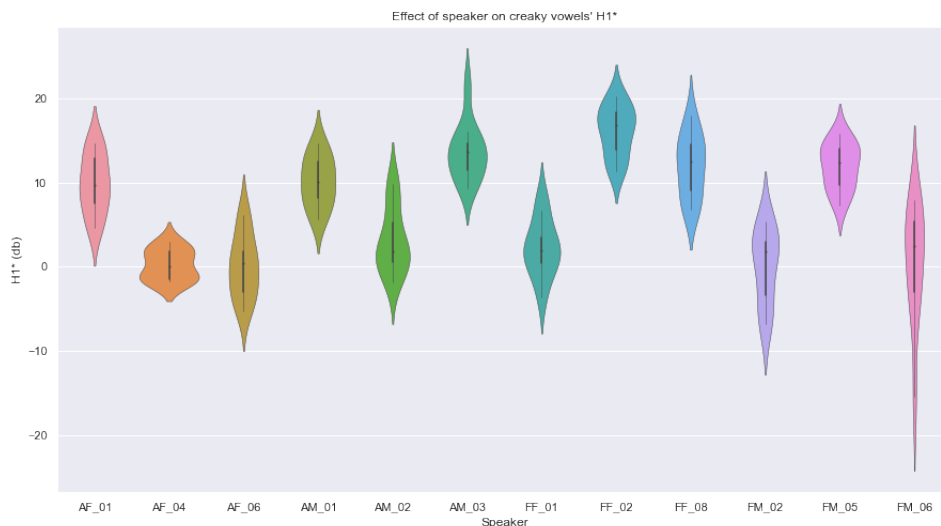


Figure 20: Violin plot representing the distribution of creaky vowels' H1* values across model speaker.

Speaker effect, illustrated in Figure 20, might be due to the fact that $H1^*$ was not normalized to the signal. A person speaking at a fixed volume at 2 cm away from the microphone will have more energy than speaking at the same volume at 4 cm from the microphone. Therefore, $H1$ (which is the magnitude of the spectrum at the first harmonic) will be larger in the 2 cm recording than in the 4 cm recording. Although we tried to keep an approximately 2 cm distance between the microphone and the mouth, subjects might have readjusted the microphone and the distance might have increased or increased, impacting $H1^*$ values. Considering that model speakers did not repeat the same sentences, hence not the same vowels, some vowels might be produced with a higher $H1^*$ than others, which would also explain Vowel effect, illustrated in Figure 21. Apart from greater variability for ROUE, there does not seem to be striking differences between vowels in terms of $H1^*$ values.

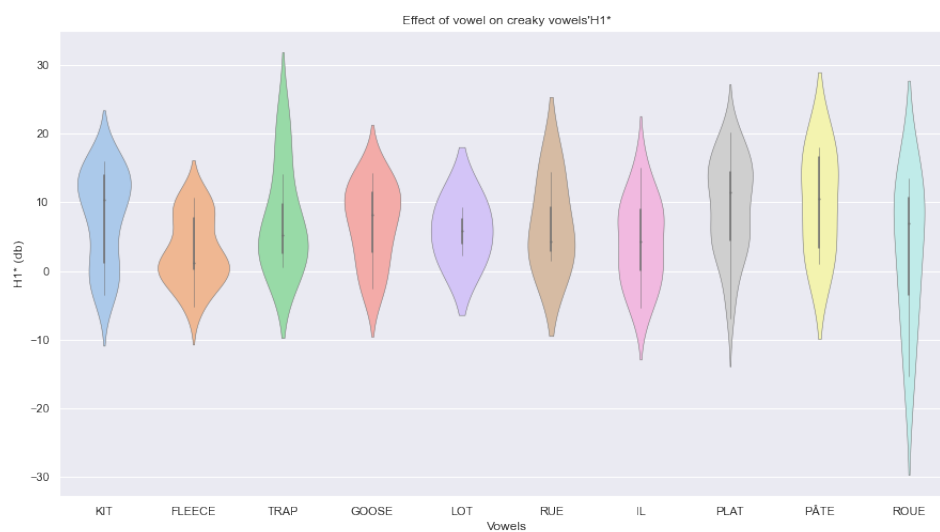


Figure 21: Violin plot representing the distribution of creaky vowels' $H1^*$ values across vowels.

Post-hoc comparisons showed that AF_06 ($M = 0.09 \pm 3.76$), FM_02 ($M = 0.1 \pm 4.51$), AF_04 ($M = 0.23 \pm 1.83$), and FM_06 ($M = 10.47 \pm 7$) produced significantly lower $H1^*$ values than AF_01 ($M = 9.84 \pm 3.54$), AM_01 ($M = 10.14 \pm 3.05$), FM_05 ($M = 11.93 \pm 2.77$), FF_08 ($M = 11.98 \pm 3.79$), (M = 141.86 \pm 34.49), AM_03 ($M = 13.69 \pm 3.46$), and

FF_02 ($M = 16.11 \pm 3.03$; $p < .001$); FF_01 ($M = 2.27 \pm 3.43$) produced significantly lower H1* values than AF_01 ($M = 9.84 \pm 3.54$), AM_01 ($M = 10.14 \pm 3.05$; $p < .01$), FM_05 ($M = 11.93 \pm 2.77$), FF_08 ($M = 11.98 \pm 3.79$), ($M = 141.86 \pm 34.49$), AM_03 ($M = 13.69 \pm 3.46$), and FF_02 ($M = 16.11 \pm 3.03$; $p < .001$); AM_02 ($M = 3.15 \pm 3.93$) produced significantly lower H1* values than AM_01 ($M = 10.14 \pm 3.05$; $p < .05$), FM_05 ($M = 11.93 \pm 2.77$), FF_08 ($M = 11.98 \pm 3.79$), ($M = 141.86 \pm 34.49$), AM_03 ($M = 13.69 \pm 3.46$), and FF_02 ($M = 16.11 \pm 3.03$; $p < .001$); and AF_01 ($M = 9.84 \pm 3.54$) produced significantly lower H1* values than FF_02 ($M = 16.11 \pm 3.03$; $p < .05$). *post-hoc* pairwise comparisons on Vowel only did not result in significant differences between vowels. Creaky vowels' H1* might only distinguish between vowels when this variable is included in a model along other variables that have a significant effect, but not on its own.

4.4 H2*

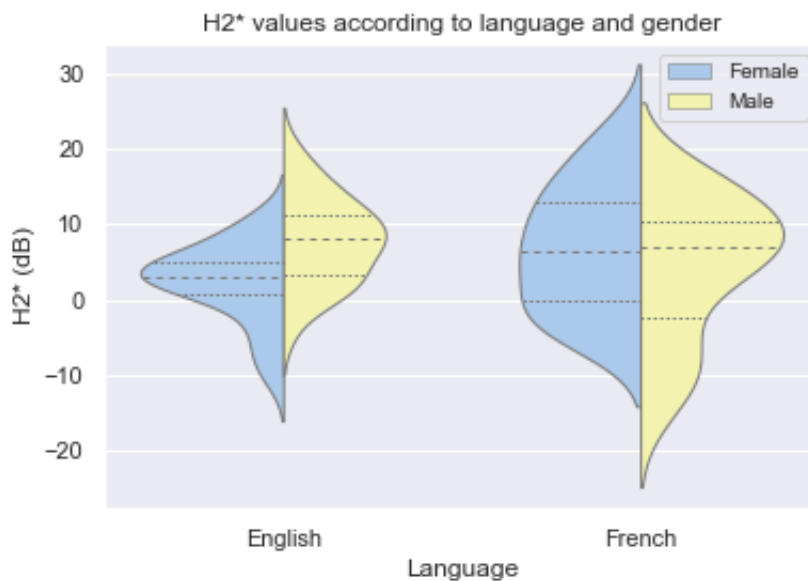


Figure 22: Violin plot representing the distribution of creaky vowels' H2* values across language and gender.

The distribution of creaky vowels' H2* values across language and gender is displayed in Figure 22. The range of H2* values is wider in French, reaching both higher and lower values than

in American English. Creaky vowels appear to be produced with a lower H2* by female American English speakers, with a peak located around 4 dB, as compared to their male counterparts for which the peak is located around 9 dB. The same pattern is observed in the production of French speakers although the differences between male and female speakers are less noticeable. There is no definite peak in the distribution of values for creaky vowels produced by female French speakers while the peak is located around 9 dB for creaky vowels produced by male French speakers, hence around the same value than the one observed in the production of American English male speakers.

Table 11: Mean H2* values (in dB) across language and gender. Standard deviation is indicated in parenthesis.

| Mean H2* and SD (dB) | | |
|------------------------|----------------|-------------|
| Language | <i>French</i> | 5.71 (8.86) |
| | <i>English</i> | 5.01 (6.12) |
| Gender | <i>Male</i> | 6.18 (7.86) |
| | <i>Female</i> | 4.68 (7.32) |
| Language:Gender | <i>FM</i> | 4.43 (9.29) |
| | <i>FF</i> | 6.91 (8.42) |
| | <i>AM</i> | 7.93 (5.75) |
| | <i>AF</i> | 2.45 (5.29) |

Mean H2* values reported in Table 11 show that creaky vowels produced by American English speakers have a lower mean H2* as compared to their French counterparts. Similarly, creaky vowels produced by female speakers have a lower mean H2* than their male counterparts. H2* is also higher for creaky vowels produced by female French speakers than when produced by male French speakers. The opposite direction is observed for American English speakers.

The model that was found for H2* was the following:

$$H2 \sim (1 \mid \text{Speaker}) + (1 \mid \text{Vowel_type})$$

Our linear-mixed effects model was fitted with H2* as dependent variable. Vowel was excluded because it did not account for enough variance. The output was the same as the one with H1*

Table 12: Significant and non-significant effects for H2* (Marginal $R^2 = 0.106$ /Conditional $R^2 = 0.6$).

| Backward reduced random-effect table: | | | |
|--|-------------------|----------------|----------------------|
| | <i>Eliminated</i> | <i>AIC</i> | <i>Pr(>Chisq)</i> |
| (1 Speaker) | 0 | 792.10 | 4.624e-10 *** |
| (1 Vowel_type) | 0 | 766.79 | 0.0002363 *** |
| Backward reduced fixed-effect table: | | | |
| | <i>Eliminated</i> | <i>F value</i> | <i>Pr(>F)</i> |
| Language:Gender | 1 | 2.6262 | 0.1319 |
| Gender | 2 | 0.2602 | 0.6196 |
| Language | 3 | 0.3729 | 0.5526 |

as dependent variable: H2* did not distinguish between neither Language, nor Gender, and the interaction between both variables was also found to be not significant. However, as can be observed in Table 12, Speaker and Vowel type were found to have a significant effect on H2*. Explanations for Speaker effect are the same than those given for H1*. Figure 23 shows the distribution of H2* values across speakers. We can observe more variability in the production of French speakers, especially in FM_02 and FM_06. FF_02 appears to produce the highest H2* values. H2* also varies more in the production of male American English speakers than in that of female American English speakers. Overall, AM_02 and FF_01 produced the lowest H2* values. *Post-hoc* comparisons showed that FF_01 ($M = -1.52 \pm 2.63$) produced significantly lower H2* values than FF_08 ($M = 7.36 \pm 5.60$; $p < .05$), AM_01 ($M = 10.12 \pm 3.62$, $p < .01$), FM_05 ($M = 11.65 \pm 5.82$), AM_03 ($M = 11.95 \pm 4.48$), and FF_02 ($M = 14.89 \pm 6.33$; $p < .001$); FM_02 ($M = -0.48 \pm 8.03$) produced significantly lower H2* values than AM_01 ($M = 10.12 \pm 3.62$, $p < .05$), FM_05 ($M = 11.65 \pm 5.82$), AM_03 ($M = 11.95 \pm 4.48$), and FF_02 ($M = 14.89 \pm 6.33$; $p < .001$); FM_06 ($M = 1.13 \pm 9.14$) produced significantly lower H2* values than AM_01 ($M = 10.12 \pm 3.62$, $p < .05$), FM_05 ($M = 11.65 \pm 5.82$; $p < .01$), AM_03 ($M = 11.95 \pm 4.48$), and FF_02 ($M = 14.89 \pm 6.33$; $p < .001$); AF_04 ($M = 1.92 \pm 5.38$) and AM_02 ($M = 2.16 \pm 3.07$) produced significantly lower H2* values than FM_05 ($M = 11.65 \pm 5.82$; $p < .05$), AM_03 ($M = 11.95 \pm 4.48$; $p < .01$), and FF_02 ($M = 14.89 \pm 6.33$; $p < .001$); FF_02 ($M = 14.89 \pm 6.33$) produced significantly higher values than AF_01 ($M = 4.27 \pm 4.08$; $p < .01$).

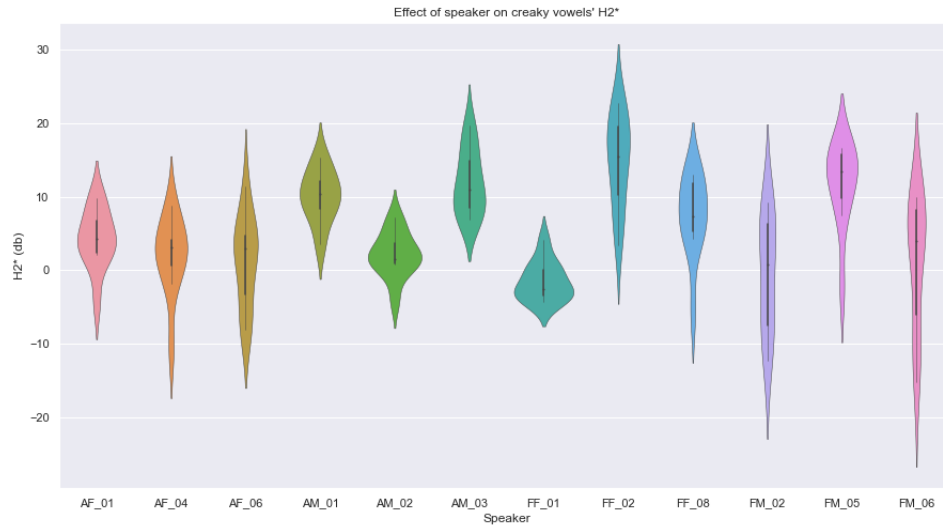


Figure 23: Violin plot representing the distribution of creaky vowels' H2* values across model speakers.

Regarding Vowel type, for which the effect is illustrated in Figure 24, *post-hoc* analysis showed that low vowels have a significantly ($p < .01$) higher H2* ($M = 8.19 \pm 8.38$) than high vowels ($M = 3.64 \pm 6.51$). This is interesting for it means that as H2* increases and gets closer to H1* (if H1* does not increase proportionally), the difference between these two harmonics decreases, hence the degree of constriction. Our conclusions might corroborate Panfili's (2015) who found that low vowels tended to be creakier than high vowels. Our following analysis on H1*-H2* will give us more reliable results.

4.5 H1*-H2*

The distribution of creaky vowels' H1*-H2* values across language and gender is displayed in Figure 25. Overall, the range of H1*-H2* values is greater in American English, reaching both higher and lower values than in French. The distribution of H1*-H2* values is much varied for female American English speakers, with no definite peak, as compared to their male counterparts, for which the peak is located around 4 dB. The distribution of values is similar across gender

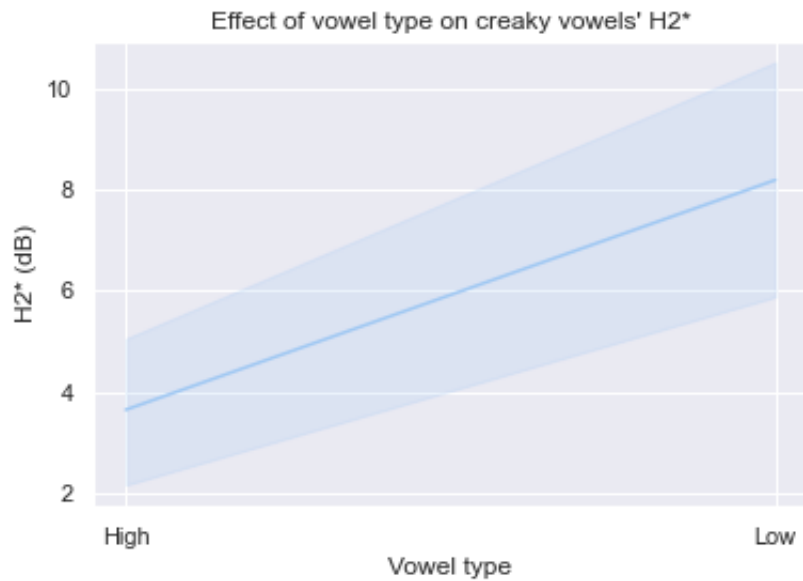


Figure 24: Effect of vowel type on creaky vowels' H2*. The shaded area represents the standard deviation.

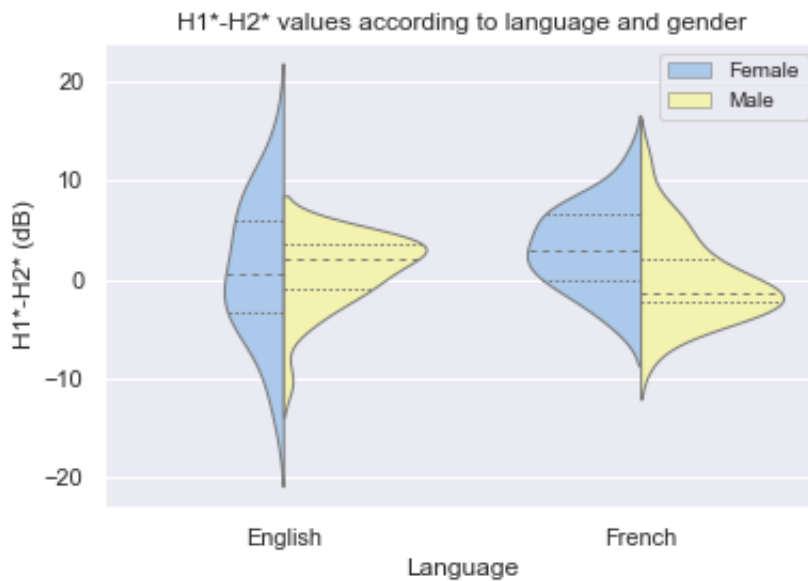


Figure 25: Violin plot representing the distribution of creaky vowels' H1*-H2* values across language and gender.

in French, with a peak located around 3 dB for female speakers and around -2 dB for male speakers. The fact that creaky vowels' H1*-H2* when produced by female American English speakers varies consequently might indicate the presence of both constricted and non-constricted

creaky voice.

Table 13: Mean H1*-H2* values (in dB) across language and gender. Standard deviation is indicated in parenthesis.

| Mean H1*-H2* and SD (dB) | | |
|--------------------------|----------------|-------------|
| Language | <i>French</i> | 1.67 (4.47) |
| | <i>English</i> | 0.96 (5.46) |
| Gender | <i>Male</i> | 0.51 (3.9) |
| | <i>Female</i> | 2.07 (5.74) |
| Language:Gender | <i>FM</i> | 0.03 (4.28) |
| | <i>FF</i> | 3.21 (4.14) |
| | <i>AM</i> | 0.98 (3.49) |
| | <i>AF</i> | 0.94 (6.87) |

As can be observed in Table 13, mean H1*-H2* is higher for American English speakers than it is for French speakers. The same observation can be made for male speakers as compared to female speakers. H1*-H2* also seems to be higher when produced by female French speakers than when produced by their male counterparts. The opposite direction is observed in American English, but the difference between male and female is very small. As we have seen previously, H1*-H2* is a correlate of glottal constriction, and creaky voice is realised with low H1*-H2* values (apart from non-constricted creaky voice). Our observations therefore suggest that creaky vowels are realised with a more constricted glottis in American English than in French. The same observation can be made for male speakers as compared to female speakers. There is no gender difference in American English in terms of degree of constriction, whereas the glottis seems to be more constricted for male French speakers than for female French speakers. This could indicate that female French speakers produce non-constricted creaky voice. We cannot corroborate observations made by Pépiot (2014) that male American English speakers have lower H1*-H2* values than male French speakers.

The linear mixed-effects model we fitted with H1*-H2* as dependent variable did not output any significant effect. We cannot conclude that H1*-H2* significantly distinguishes creaky vowels between language and gender. The degree of constriction of creaky vowels, therefore, does not significantly vary among the different groups observed.

4.6 H1*-A1*

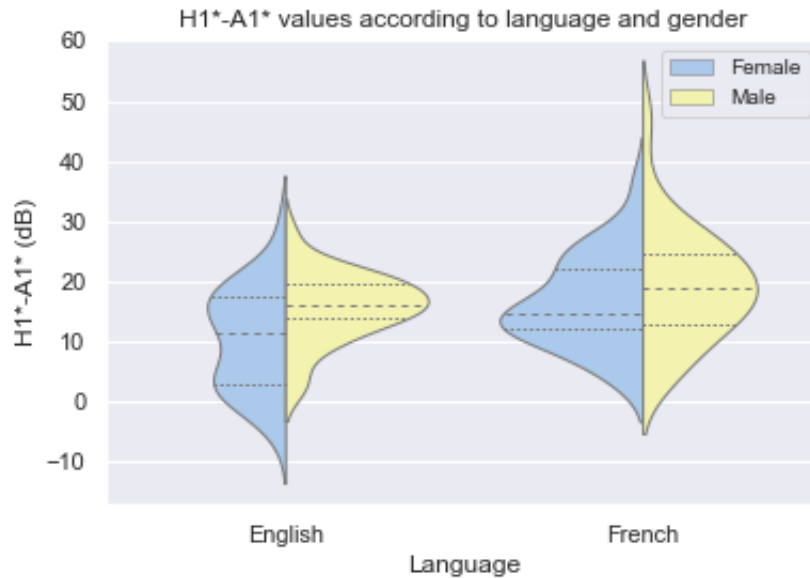


Figure 26: Violin plot representing the distribution of creaky vowels' H1*-A1* values across language and gender.

The distribution of creaky vowels' H1*-A1* values across language and gender is displayed in Figure 26. Overall, H1*-A1* values seem to be concentrated in two similar areas in both languages, although the minima seem lower in American English and the maxima higher in French. Two distinct peaks can be observed in the distribution of H1*-A1* values for female American English speakers, one around 3 dB and the other around 17 dB. For male American speakers, the peak is located around 18 dB. The distribution of H1*-A1* values is also quite similar for creaky vowels produced by French speakers, with a peak located around 14 dB for female speakers and around 20 dB for male speakers.

Mean H1*-A1* values reported in Table 14 show that creaky vowels are produced with a higher H1*-A1* by French speakers as compared to when produced by American English speakers, and with a higher H1*-A1* when produced by male speakers as compared to when produced by female speakers. H1*-A1* is also higher for creaky vowels produced by male French speakers than by their female counterparts. The same pattern is observed in American English. A high

Table 14: Mean H1*-A1* values (in dB) across language and gender. Standard deviation is indicated in parenthesis.

| Mean H1*-A1* and SD (dB) | | |
|--------------------------|----------------|--------------|
| Language | <i>French</i> | 17.94 (8.5) |
| | <i>English</i> | 12.98 (7.59) |
| Gender | <i>Male</i> | 17.55 (7.85) |
| | <i>Female</i> | 13.51 (8.49) |
| Language:Gender | <i>FM</i> | 19.35 (9.36) |
| | <i>FF</i> | 16.63 (7.55) |
| | <i>AM</i> | 15.75 (5.6) |
| | <i>AF</i> | 10.4 (8.35) |

H1*-A1* has been found to correlate with the presence of a posterior glottal chink (Hanson & Chuang, 1999), suggesting less high frequency energy and weaker F1 peak. Our data suggest that creaky vowels might therefore be produced with a larger posterior glottal chink in French and by male speakers. We cannot corroborate the fact that the presence of a posterior glottal chink persisting through the entire glottal cycle is more common for female than for male speakers (Hanson & Chuang, 1999).

We fitted our linear mixed-effects model with H1*-A1* as dependent variable. Results, reported in Table 15, show that H1*-A1* did not distinguish between Language and Gender. The interaction between both variables was also found to be not significant. However, both Speaker and Vowel displayed significant effects. They are illustrated in Figure 27 and 28, respectively.

The model that was found for H1*-A1* was the following:

$$\text{H1A1} \sim (1 \mid \text{Speaker}) + (1 \mid \text{Vowel})$$

Some speakers might realise creaky vowels with a larger posterior glottal opening than others, indicating they might have produced occurrences of non-constricted creak. There is more observed inter-individual variability in the production of FF_01 and FM_02, for instance. AF_04 seems to produce the lowest H1*-A1* values. *Post-hoc* comparisons showed that AF_04 ($M = 5.13 \pm 6.31$) produced lower H1*-A1* values than AM_02 ($M = 16.56 \pm 6.16$; $p < .05$), FF_01 ($M = 19.29 \pm 9.86$; $p < .01$), FM_05 ($M = 20.67 \pm 6.56$) and FM_02 ($M = 22.72 \pm 13.94$; p

Table 15: Significant and non-significant effects for H1*-A1* (Marginal $R^2 = 0.071$ /Conditional $R^2 = 0.556$).

| Backward reduced random-effect table: | | | |
|--|-------------------|------------|----------------------|
| | <i>Eliminated</i> | <i>AIC</i> | <i>Pr(>Chisq)</i> |
| (1 Vowel_type) | 1 | 782.73 | 0.1095075 |
| (1 Speaker) | 0 | 795.10 | 0.0001504 *** |
| (1 Vowel) | 0 | 814.36 | 6.652e-09 *** |

| Backward reduced fixed-effect table: | | | |
|---|-------------------|----------------|------------------|
| | <i>Eliminated</i> | <i>F value</i> | <i>Pr(>F)</i> |
| Language:Gender | 1 | 0.6319 | 0.4435 |
| Gender | 2 | 0.3127 | 0.5847 |
| Language | 3 | 2.3761 | 0.1516 |

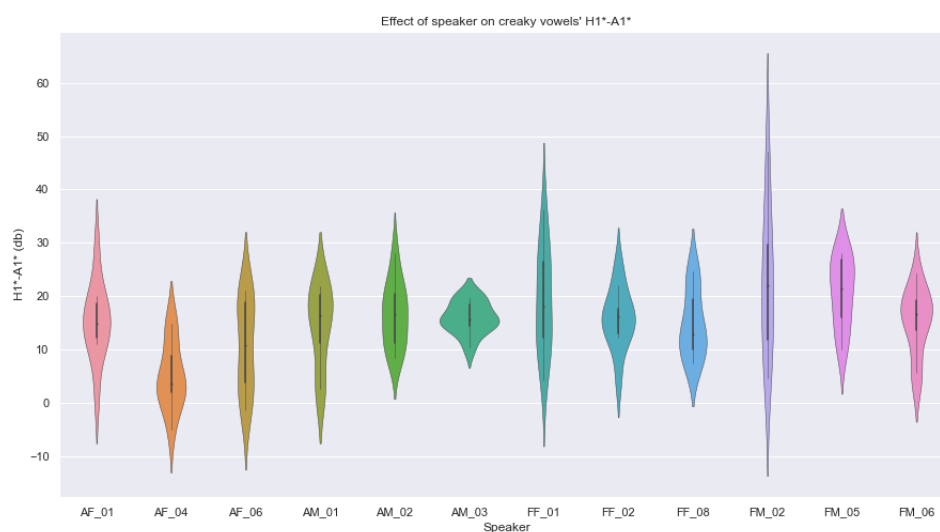


Figure 27: Violin plot representing the distribution of creaky vowels' H1*-A1* across model speakers.

<.001). We have to bear in mind that vowels differ across model speakers, which might account for the Vowel effect observed.

Post-hoc comparisons showed that PÂTE ($M = 20.64 \pm 3.86$) was produced with significantly lower H1*-A1* values than FLEECE ($M = 7.98 \pm 6.16$; $p < .05$); and PLAT ($M = 21.76 \pm 6.23$)

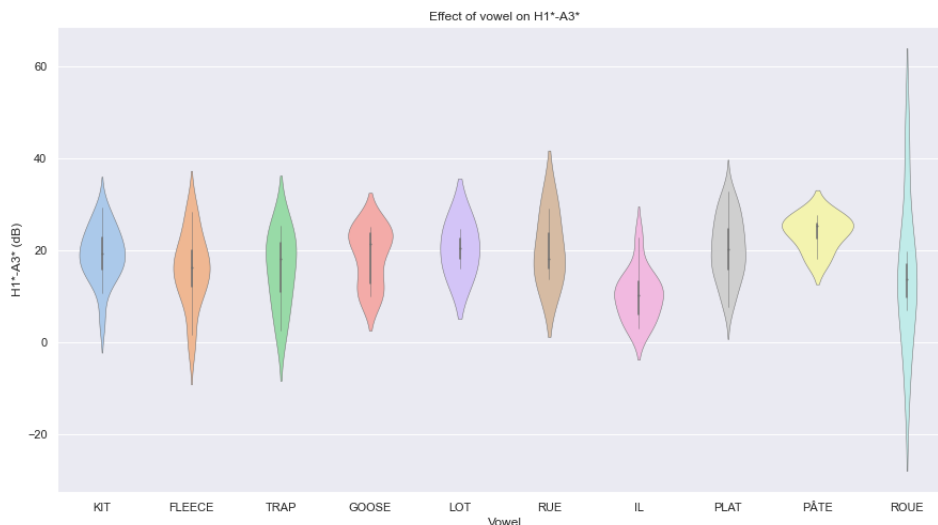


Figure 28: Violin plot representing the distribution of creaky vowels' H1*-A1* across vowels.

was produced with significantly lower H1*-A1* values than FLEECE ($M = 7.98 \pm 6.16$) and IL ($M = 9.29 \pm 3.6$; $p < .05$). Vowel effect might then be related to Vowel type, but the uneven distribution of low and high vowels in our dataset ($N_{High} = 71$; $N_{Low} = 45$) could account for the fact that there is a Vowel effect but not a Vowel type effect. As mentioned above, weaker F1 peaks are correlated with high H1*-A1* values. Considering that low vowels have high F1 peaks, it might be more articulatory constraining to produce creaky voice on high vowels.

4.7 H1*-A2*

The distribution of creaky vowels' H1*-A2* values across language and gender is displayed in Figure 29. Overall, the distribution of H1*-A2* values appear similar across language but with higher maxima in French and lower minima in American English. Male American English speakers seem to produce creaky vowels with overall higher H1*-A2* values, with a peak located around 22 dB, than their female counterparts for which the peak is located around 17 dB. The distribution of H1*-A2* values is greater for female American English speakers as compared to that of their male counterparts. It reaches lower and negative values. The distribution of

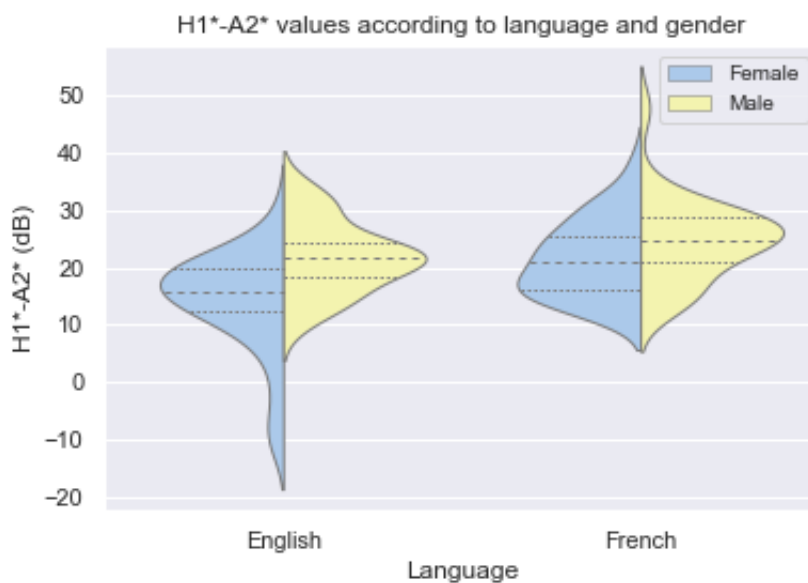


Figure 29: Violin plot representing the distribution of creaky vowels' H1*-A2* values across language and gender.

H1*-A2* values is very similar across gender in French, although male speakers tend to produce creaky vowels with higher H1*-A2* values, with a peak located around 27 dB, than their female counterparts for which the peak is located around 18 dB.

Table 16: Mean H1*-A2* values (in dB) across language and gender. Standard deviation is indicated in parenthesis.

| Mean H1*-A2* and SD (dB) | | |
|--------------------------|----------------|--------------|
| Language | <i>French</i> | 23.09 (7.01) |
| | <i>English</i> | 17.83 (8.45) |
| Gender | <i>Male</i> | 23.28 (6.76) |
| | <i>Female</i> | 17.83 (8.54) |
| Language:Gender | <i>FM</i> | 24.75 (7.13) |
| | <i>FF</i> | 21.54 (6.63) |
| | <i>AM</i> | 21.81 (6.14) |
| | <i>AF</i> | 14.11 (8.7) |

Mean H1*-A2* values reported in Table 16 show that, overall, creaky vowels are produced with a higher mean H1*-A2* by French speakers as compared to American English speakers, and with a higher H1*-A2* when produced by male speakers than by female speakers. H1*-A2* is

also higher for creaky vowels produced by male French speakers as compared to when produced by their female counterparts. The same pattern is observed in American English. creaky voice is characterised by more skewing and abrupt changes in the shape of the glottal pulse, which is a consequence of airflow building up gradually as the vocal folds open, then dropping suddenly when they close abruptly, resulting in low H1*-A2* values. Our observations therefore suggest that creaky vowels are characterised by more skewing and abrupt changes in the shape of the glottal pulse in American English than in French, and when produced by female than by male speakers.

We ran a linear mixed-effects model with H1*-A2* as dependent variable. Results reported in Table 17 show that, as for H1*-A1*, H1*-A2* did not distinguish creaky vowels between neither Language, nor Gender. The interaction between both variables was also found to be not significant. However, both Speaker and Vowel displayed significant effects. They are illustrated in Figure 30 and 31, respectively.

Table 17: Significant and non-significant effects for H1*-A2* (Marginal R² = 0.155/Conditional R² = 0.476).

| Backward reduced random-effect table: | | | |
|--|-------------------|----------------|----------------------|
| | <i>Eliminated</i> | <i>AIC</i> | <i>Pr(>Chisq)</i> |
| (1 Vowel_type) | 1 | 782.51 | 0.7160366 |
| (1 Speaker) | 0 | 790.80 | 0.0013431 ** |
| (1 Vowel) | 0 | 793.85 | 0.0002601 *** |
| Backward reduced fixed-effect table: | | | |
| | <i>Eliminated</i> | <i>F value</i> | <i>Pr(>F)</i> |
| Language:Gender | 1 | 0.9927 | 0.33972 |
| Gender | 2 | 1.3769 | 0.26115 |
| Language | 3 | 4.2574 | 0.06335 |

The model that was found for H1*-A2* was the following:

$$H1A2 \sim (1 | Speaker) + (1 | Vowel)$$

Some speakers might realise creaky vowels with more abrupt vocal fold closure than others. AF_06 displays the lowest H1*-A2* values, indicating more abrupt vocal fold closure, while FM_02 displays the highest H1*-A2* values, indicating less abrupt vocal fold closure. There is particularly less inter-variability observed in the production of AF_04, AM_01, and FM_05. *Post-hoc* comparisons showed that AF_06 ($M = 7.98 \pm 11.41$) produced lower H1*-A2* values than AM_03 ($M = 18.86 \pm 19.18$), FF_02 ($M = 19.18 \pm 4.28$; $p < .05$), FF_08 ($M = 21.12 \pm 6.96$; $p < .01$), FM_06 ($M = 22.61 \pm 4.82$), AM_01 ($M = 23.39 \pm 4.19$), AM_02 ($M = 23.5 \pm 7.56$), FM_02 ($M = 23.52 \pm 11.8$), FF_01 ($M = 24.32 \pm 7.74$), and FM_05 ($M = 27.87 \pm 1.93$; $p < .001$).

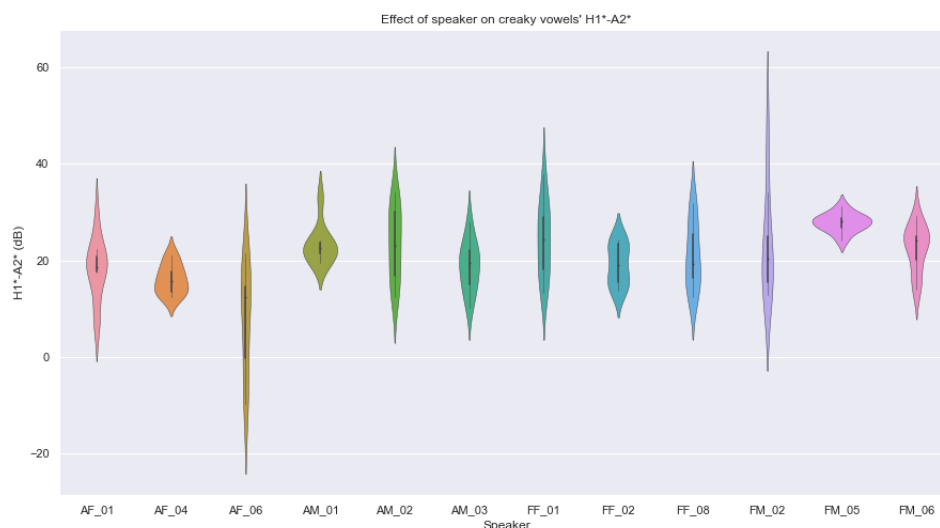


Figure 30: Violin plot representing the distribution of creaky vowels' H1*-A2* across model speakers.

Results also show us that some vowels seem to be realised with more abrupt glottal fold closure than others. We can observe that FLEECE displays the lowest H1*-A2* values. There is more variability for LOT and ROUE. *Post-hoc* comparisons confirmed that FLEECE ($M = 13.18 \pm 9.45$) was produced with significantly lower H1*-A2* values than ROUE ($M = 24.91 \pm 11.16$), PÂTE ($M = 26.49 \pm 4.37$; $p < .05$), GOOSE ($M = 24.67 \pm 6.03$; $p < .05$), PLAT ($M = 24.21 \pm 5.96$; $p < .001$); and PLAT ($M = 24.21 \pm 5.96$) was produced with significantly higher values than KIT ($M = 17.12 \pm 5.83$; $p < .001$).

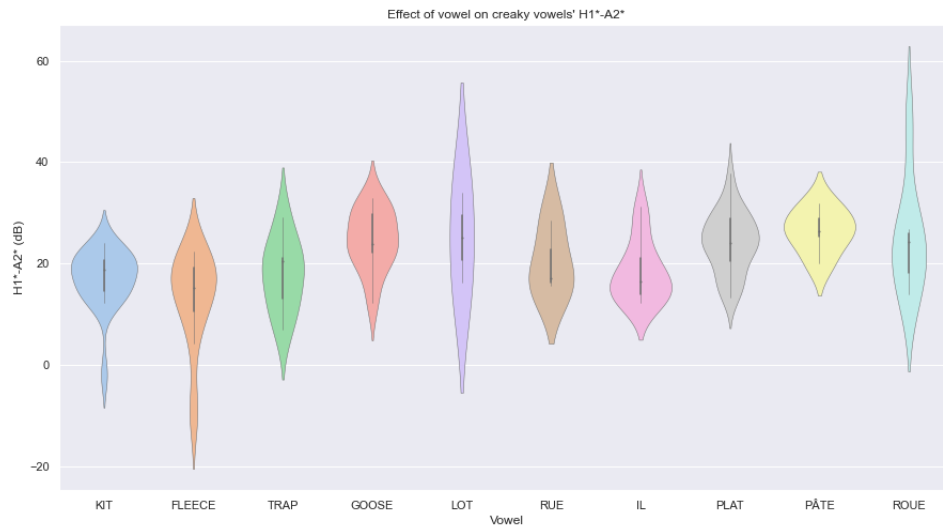


Figure 31: Violin plot representing the distribution of creaky vowels' H1*-A2* across vowels.

4.8 H1*-A3*

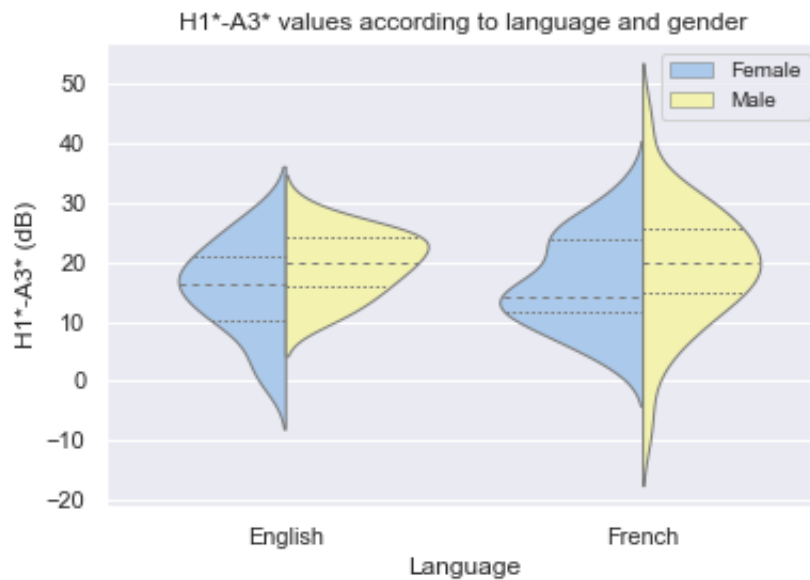


Figure 32: Violin plot representing the distribution of creaky vowels' H1*-A3* values across language and gender.

The distribution of creaky vowels' H1*-A3* values across language and gender is displayed

in Figure 32. Overall, there is more variability in the distribution of H1*-A3* values in French than there is in American English. The distribution of H1*-A3* values across gender is relatively similar in American English, with female speakers producing creaky vowels with a lower H1*-A3* than their male counterparts. The peak is located around 17 dB for female American English speakers and around 23 dB for male American English speakers. The range of H1*-A3* values is greater for creaky vowels produced by male French speakers, with a peak located around 20 dB, while it is located around 13 dB when produced by their female counterparts.

Table 18: Mean H1*-A3* values (in dB) across language and gender. Standard deviation is indicated in parenthesis.

| Mean H1*-A3* and SD (dB) | | |
|--------------------------|----------------|---------------|
| Language | <i>French</i> | 17.96 (8.88) |
| | <i>English</i> | 17.71 (6.84) |
| Gender | <i>Male</i> | 19.57 (7.94) |
| | <i>Female</i> | 16.21 (7.56) |
| Language:Gender | <i>FM</i> | 19.35 (10.02) |
| | <i>FF</i> | 16.66 (7.61) |
| | <i>AM</i> | 19.8 (5.28) |
| | <i>AF</i> | 15.76 (7.61) |

Table 18 shows that, overall, creaky vowels when produced by French speakers have a slightly higher mean H1*-A3* than when produced by American English speakers, and male speakers produce creaky vowels with a higher H1*-A3* than female speakers. H1*-A3* is also higher for creaky vowels produced by male speakers within-language. Considering that H1*-A3* is correlated with the ratio of the duration of the closed phase to the duration of a complete glottal cycle, and that the closed phase is longer in creaky phonation, resulting in lower H1*-A3* values, our data suggest that the closed phase is longer in creaky vowels produced by American English and by female speakers, than by French and male speakers. The closed phase is also longer for female than for male within-language.

The model that was found for H1*-A3* was the following:

$$\text{H1A3} \sim (1 \mid \text{Speaker}) + (1 \mid \text{Vowel})$$

Table 19: Significant and non-significant effects for H1*-A3* (Marginal $R^2 = 0.042$ /Conditional $R^2 = 0.408$).

| Backward reduced random-effect table: | | | |
|--|-------------------|----------------|----------------------|
| | <i>Eliminated</i> | <i>AIC</i> | <i>Pr(>Chisq)</i> |
| (1 Vowel_type) | 1 | 798.69 | 0.4887635 |
| (1 Speaker) | 0 | 808.99 | 0.0004537 *** |
| (1 Vowel) | 0 | 809.71 | 0.0003079 *** |
| Backward reduced fixed-effect table: | | | |
| | <i>Eliminated</i> | <i>F value</i> | <i>Pr(>F)</i> |
| Language:Gender | 1 | 0.2180 | 0.6493 |
| Gender | 2 | 0.0292 | 0.8667 |
| Language | 3 | 1.5239 | 0.2417 |

We ran a linear mixed-effects model with H1*-A3* as dependent variable. Results reported in Table 19, show that, as for H1*-A1* and H1*-A2*, neither Language, nor Gender, had a significant effect on creaky vowels' H1*-A3*. The interaction between the two variables appeared not to be significant either. However, both Speaker and Vowel displayed significant effects. They are both illustrated in Figure 33 and 34, respectively.

Some speakers might realise creaky vowels with longer closed phase than others. What we can observe is that there seems to be slightly more variability in the production of French speakers. *Post-hoc* comparisons showed that AF_06 ($M = 10.96 \pm 7.51$) produced significantly lower H1*-A3* values than FM_05 ($M = 24.67 \pm 5.95$; $p < .01$). As for speaker effect, some vowels might be realised with longer closed phase than others. ROUE displays a great deal of variability while IL seems to display the lowest H1*-A3* values. *Post-hoc* comparisons showed that IL ($M = 10.27 \pm 5.52$) was produced with significantly lower H1*-A3* values than KIT ($M = 19.12 \pm 5.96$), PÂTE ($M = 23.84 \pm 3.73$; $p < .05$), and PLAT ($M = 20.47 \pm 6.9$; $p < .01$).

4.9 H2*-H4*

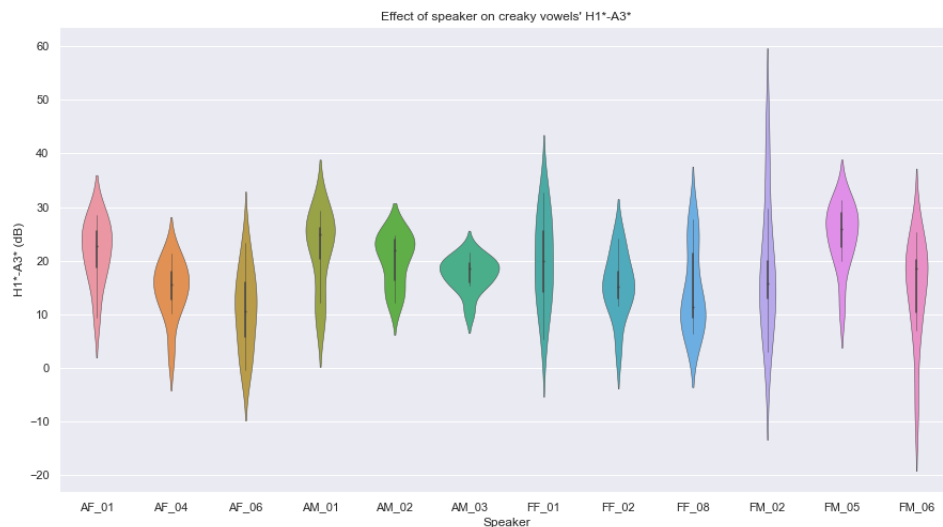


Figure 33: Violin plot representing the distribution of creaky vowels' H1*-A3* values across model speakers.

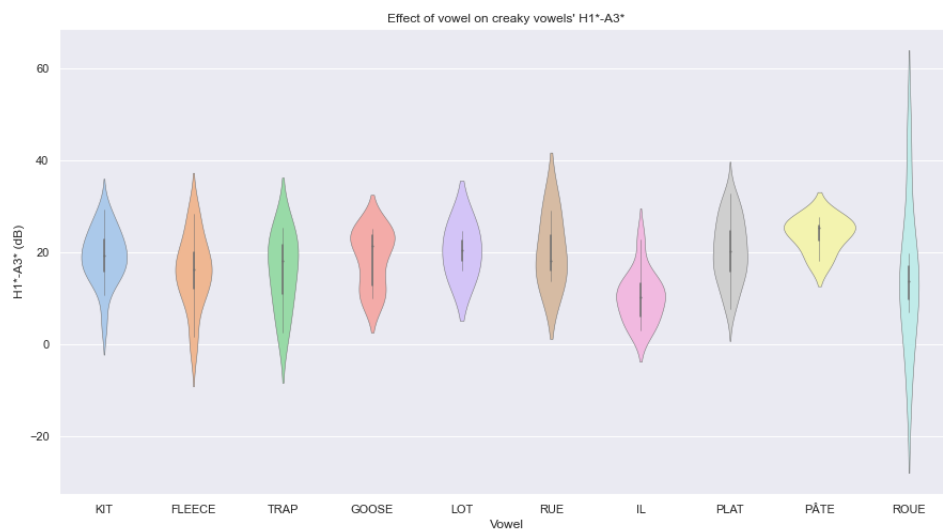


Figure 34: Violin plot representing the distribution of H1*-A3* values across vowels.

The distribution of creaky vowels' H2*-H4* values across language and gender is displayed in Figure 35. The distribution of H2*-H4* values follow the same pattern across language, although there appears to be slightly more variability for French speakers than there is for American

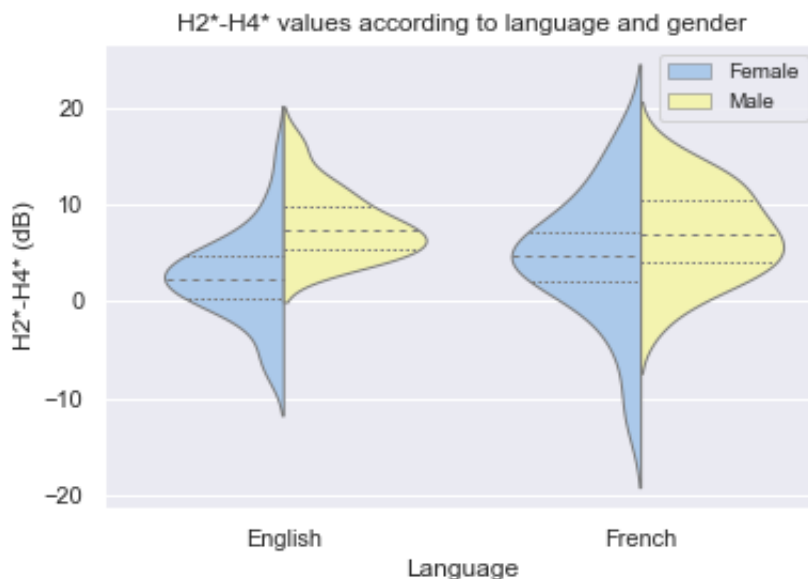


Figure 35: Violin plot representing the distribution of creaky vowels' $H2^*-H4^*$ values across language and gender.

English speakers. Male American English speakers tend to produce creaky vowels with a higher $H2^*-H4^*$, with a peak located around 7 dB, than their female counterparts for which the peak is located around 2 dB. For French speakers the peak is located around 4 dB for female speakers and around 5 dB for male speakers. There is also more variability in the distribution of female speakers in both languages, in which the minima is lower than that of their male counterparts, and lower for French female than for their American English counterparts.

Table 20: Mean $H2^*-H4^*$ values (in dB) across language and gender. Standard deviation is indicated in parenthesis.

| Mean $H2^*-H4^*$ and SD (dB) | | |
|------------------------------|----------------|-------------|
| Language | <i>French</i> | 5.79 (4.09) |
| | <i>English</i> | 5.24 (4.92) |
| Gender | <i>Male</i> | 7.53 (4.09) |
| | <i>Female</i> | 3.63 (5.86) |
| Language:Gender | <i>FM</i> | 7.04 (4.64) |
| | <i>FF</i> | 4.63 (6.78) |
| | <i>AM</i> | 8.08 (3.47) |
| | <i>AF</i> | 2.64 (4.68) |

We can observe in Table 20 below that, overall, creaky vowels when produced by French speakers have a slightly higher H2*-H4* as compared to when produced by American English speakers, and male speakers produce creaky vowels with a higher H2*-H4* than female speakers. H2*-H4* is also higher for creaky vowels produced by male speakers within-language.

Table 21: Significant and non-significant effects for H2*-H4* (Marginal R² = 0.157/Conditional R² = 0.235).

| Backward reduced random-effect table: | | | |
|--|-------------------|----------------|----------------------|
| | <i>Eliminated</i> | <i>AIC</i> | <i>Pr(>Chisq)</i> |
| (1 Speaker) | 1 | 710.44 | 0.62279 |
| (1 Vowel_type) | 0 | 711.37 | 0.08654 |
| Backward reduced fixed-effect table: | | | |
| | <i>Eliminated</i> | <i>F value</i> | <i>Pr(>F)</i> |
| Language:Gender | 1 | 3.2629 | 0.0735 |
| Language | 2 | 0.1786 | 0.6735 |
| Gender | 0 | 18.4312 | 3.716e-05 *** |

The model that was found for H2*-H4* was the following:

$$\text{H2H4} \sim \text{Gender} + (1 \mid \text{Vowel_type})$$

We fitted our linear mixed-effects model with H2*-H4* as dependent variable. Vowel was excluded because it did not account for enough variance. Results reported in Table 21 show that only Gender had a statistically significant effect on creaky vowels' H2*-H4*. The effect is displayed in Figure 36. We can conclude that creaky vowels produced by female speakers have a significantly lower H2*-H4*, which might indicate a creakier type of phonation, as has previously been reported (Garellek & Seyfarth, 2016).

4.10 CPP

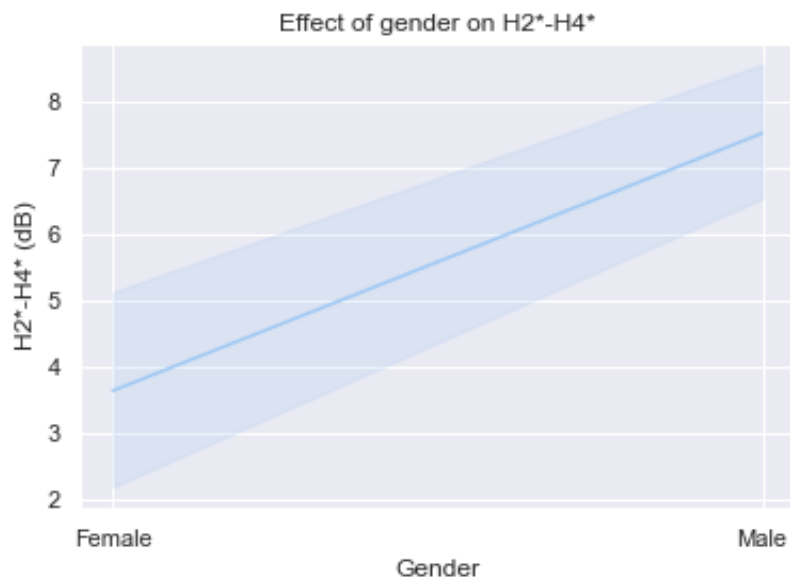


Figure 36: Effect of gender on creaky vowels' H2*-H4*. The shaded area represents the standard deviation.

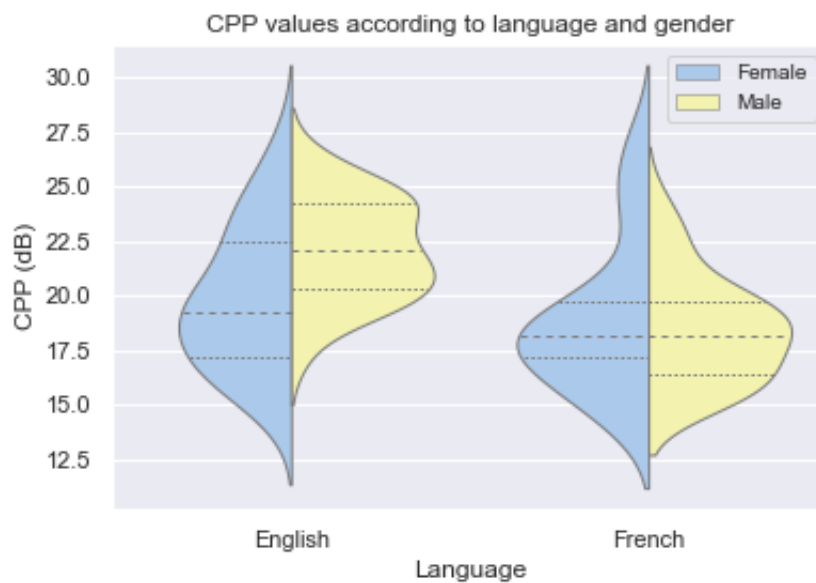


Figure 37: Violin plot representing the distribution of creaky vowels' CPP values across language and gender.

The distribution of creaky vowels' CPP values across language and gender is displayed in Figure 37. Overall, the range of CPP values appears to be similar across language. Male American English speakers produce creaky vowels with a higher CPP, with a more prominent

peak around 21 dB, than their female counterparts for which the peak is located around 18.5 dB. CPP does not seem to vary much for creaky vowels when produced by both male and female French speakers. The peak is located around 18 dB for female French speakers and around 18.5 dB for male French speakers. Both French and American English female speakers have a wider range of values than their male counterparts, with both higher maxima and lower minima values.

Table 22: Mean CPP values (in dB) across language and gender. Standard deviation is indicated in parenthesis.

| Mean CPP and SD (dB) | | |
|------------------------|----------------|--------------|
| Language | <i>French</i> | 18.71 (2.99) |
| | <i>English</i> | 20.98 (3.01) |
| Gender | <i>Male</i> | 20.28 (2.98) |
| | <i>Female</i> | 19.44 (3.36) |
| Language:Gender | <i>FM</i> | 18.44 (2.43) |
| | <i>FF</i> | 18.96 (3.45) |
| | <i>AM</i> | 22.12 (2.28) |
| | <i>AF</i> | 19.91 (3.25) |

Results, reported in Table 22 show that, overall, creaky vowels when produced by American English speakers have a higher CPP as compared to when produced by French speakers. The same observation can be made for male speakers as compared to female speakers. CPP is also higher for creaky vowels produced by female French speakers as compared to when produced by male French speakers (although the difference is small). The opposite direction is observed for American English speakers. CPP is a measure of periodicity: the greater the difference, the greater the ratio of periodic to aperiodic sound in the signal. Our data suggest that there is more aperiodicity in creaky vowels produced by French speakers than in those produced by American English speakers. The same pattern is observed in the production of female speakers as compared to that of male speakers.

We ran a linear mixed-effects model with CPP as dependent variable. Results, reported in Table 23, show that Language, Speaker and Vowel type had a statistically significant effect on creaky vowels' CPP.

Table 23: Significant and non-significant effects for CPP (Marginal $R^2 = 0.265$ /Conditional $R^2 = 0.544$).

| Backward reduced random-effect table: | | | |
|--|-------------------|----------------|----------------------|
| | <i>Eliminated</i> | <i>AIC</i> | <i>Pr(>Chisq)</i> |
| (1 Vowel) | 1 | 559.23 | 0.349762 |
| (1 Speaker) | 0 | 577.02 | 8.647e-06 *** |
| (1 Vowel_type) | 0 | 565.82 | 0.003374 ** |
| Backward reduced fixed-effect table: | | | |
| | <i>Eliminated</i> | <i>F value</i> | <i>Pr(>F)</i> |
| Language:Gender | 1 | 2.4798 | 0.14207 |
| Gender | 2 | 0.6910 | 0.42250 |
| Language | 0 | 7.9281 | 0.01486 * |

The model that was found for CPP was the following:

$$\text{CPP} \sim \text{Language} + (1 | \text{Speaker}) + (1 | \text{Vowel_type})$$

The distribution of CPP values across language is displayed in Figure 38. We can observe more aperiodicity in creaky vowels produced by French speakers. We should, however, be careful in our interpretation here. Low CPP values can also indicate a breathier type of voice quality, and not all types of creaky voice have low CPP values. It might be that creaky vowels were actually realised with more air leakage by French speakers than by American English speakers, or that they used more vocal fry (which has low CPP values).

Figure 39 shows the distribution of CPP values across speakers. This violin plot confirms that French speakers produce creaky vowels with lower CPP values than American English speakers. AF_04 seems to produce creaky voice similarly as French speakers in terms of CPP, with lower values, hence with more aperiodicity. There is also more inter-speaker variability in American English than in French, and more intra-speaker variability for AF_06, AM_03, FF_02, FM_06.

Post-hoc comparisons showed that AF_04 ($M = 17 \pm 1.43$) produced significantly lower CPP values than AF_06 ($M = 21.24 \pm 3.83$), AM_03 ($M = 21.43 \pm 2.99$), AF_01 ($M = 21.48$

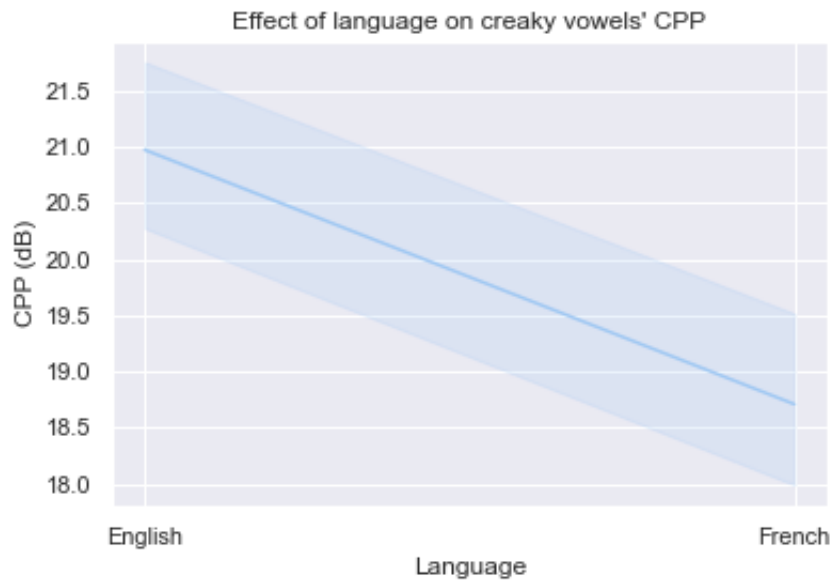


Figure 38: Effect of language on creaky vowels' CPP. The shaded area represents the standard deviation.

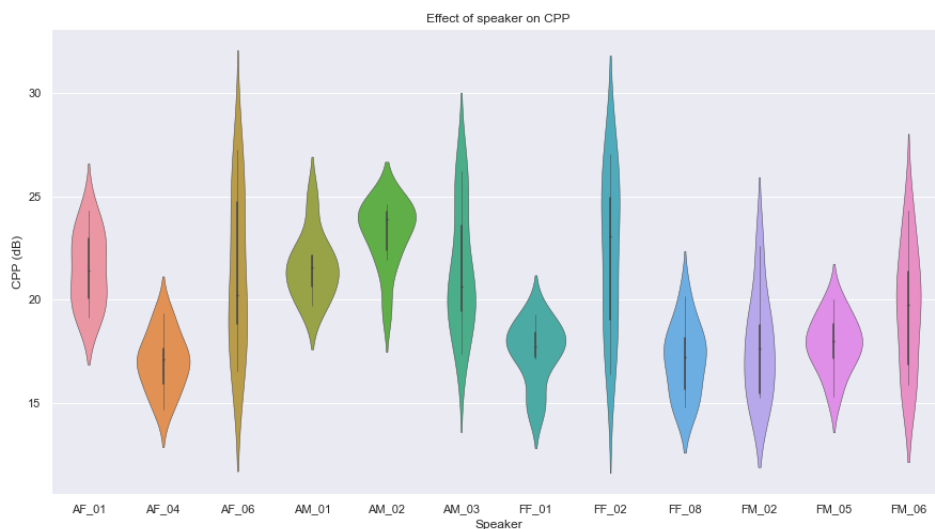


Figure 39: Violin plot representing the distribution of CPP values across model speakers.

± 1.81), AM_01 ($M = 21.62 \pm 1.58$; $p < .01$), FF_02 ($M = 22.31 \pm 3.78$), and AM_02 ($M = 23.21 \pm 1.61$; $p < .001$); FF_08 ($M = 17.12 \pm 1.73$) produced significantly lower CPP values than AF_06 ($M = 21.24 \pm 3.83$; $p < .05$), AM_03 ($M = 21.43 \pm 2.99$), AF_01 ($M = 21.48 \pm 1.81$), AM_01 ($M = 21.62 \pm 1.58$; $p < .01$), FF_02 ($M = 22.31 \pm 3.78$), and AM_02 ($M =$

23.21 ± 1.61 ; $p < .001$); FF_01 ($M = 17.47 \pm 1.49$) produced significantly lower CPP values than AF_06 ($M = 21.24 \pm 3.83$), AM_03 ($M = 21.43 \pm 2.99$), AF_01 ($M = 21.48 \pm 1.81$), AM_01 ($M = 21.62 \pm 1.58$; $p < .05$), FF_02 ($M = 22.31 \pm 3.78$; $p < .01$), and AM_02 ($M = 23.21 \pm 1.61$; $p < .001$); FM_02 ($M = 17.74 \pm 2.53$) produced significantly lower CPP values than FF_02 ($M = 22.31 \pm 3.78$; $p < .01$) and AM_02 ($M = 23.21 \pm 1.61$; $p < .001$); FM_05 ($M = 17.86 \pm 1.37$) produced significantly lower values than FF_02 ($M = 22.31 \pm 3.78$; $p < .01$) and AM_02 ($M = 23.21 \pm 1.61$; $p < .001$).

Figure 40 shows the effect of Vowel type in CPP values. High vowels seem to produce with lower CPP values ($M = 19.63 \pm 3.27$) than low vowels ($M = 20.18 \pm 3.08$), which means that more aperiodicity is observed in high vowels than in low vowels in our dataset. *Post-hoc* pairwise comparisons on Vowel type only did not result in significant difference between low and high vowels. Creaky vowels' CPP might only distinguish between vowel types when this variable is included in a model along other variables that have a significant effect, but not on its own.

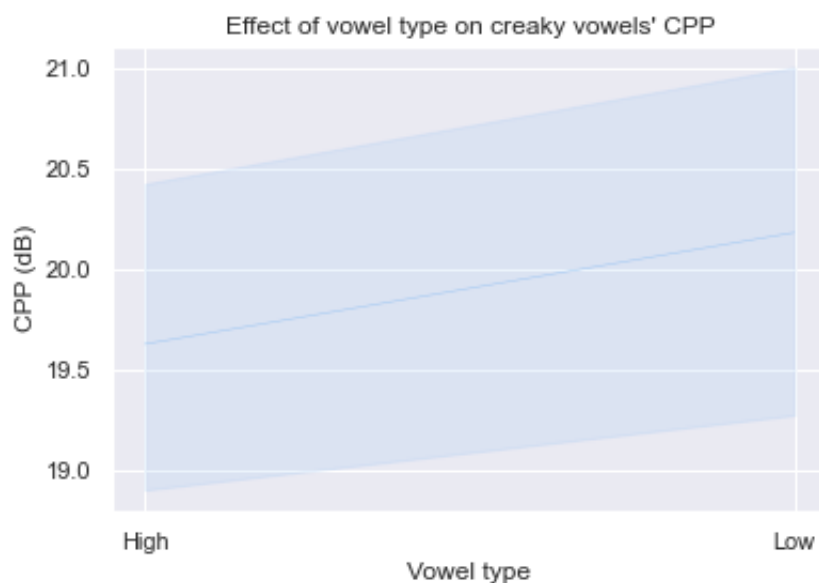


Figure 40: Effect of vowel type on creaky vowels' CPP. The shaded area represents the standard deviation.

In this section we analysed all acoustic measures. Electroglottographic analyses are reported in the following section.

5 Analysis of EGG measures

In this section we provide an analysis of electroglottographic measures. Analyses of CQ and PIC are reported in section 5.1 and 5.2, respectively.

5.0.1 Closed quotient

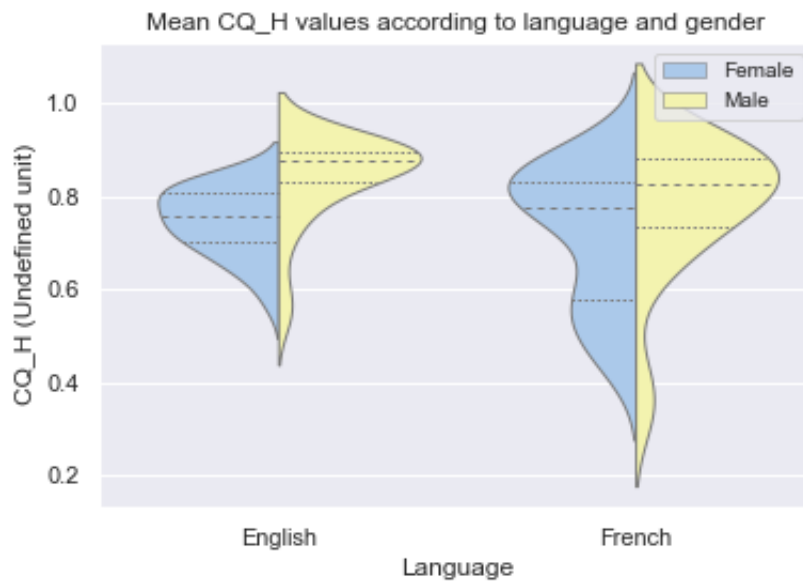


Figure 41: Violin plot representing the distribution of creaky vowels' CQ values across language and gender.

The distribution of creaky vowels' CQ values across language and gender is displayed in Figure 41. Overall, the range of CQ values appears to be greater for French speakers than it is for American English speakers, reaching lower values. Male American English speakers produce creaky vowels with a higher CQ, with a more prominent peak around 0.9, than their female counterparts for which the peak is located around 0.79. The distribution of CQ values is quite similar across gender in French, with a peak located around 0.84 for both male and female speakers.

Table 24 shows that, overall, creaky vowels when produced by American English speakers have a slightly higher CQ than when produced by French speakers. The same observation can

Table 24: Mean CQ values across language and gender. Standard deviation is indicated in parenthesis

| Mean CQ and SD (undefined unit) | | |
|---------------------------------|----------------|-------------|
| Language | <i>French</i> | 0.74 (0.15) |
| | <i>English</i> | 0.79 (0.1) |
| Gender | <i>Male</i> | 0.81 (0.13) |
| | <i>Female</i> | 0.73 (0.12) |
| Language:Gender | <i>FM</i> | 0.78 (0.15) |
| | <i>FF</i> | 0.72 (0.15) |
| | <i>AM</i> | 0.84 (0.1) |
| | <i>AF</i> | 0.75 (0.07) |

be made for male speakers as compared to female speakers. CQ is also higher for creaky vowels produced by male French speakers as compared to when produced by female French speakers. The pattern is the same for American English speakers. creaky voice is correlated with a low OQ (and low H1-H2), hence with a high CQ. Our observations suggest that the glottis spends more time closed when creaky vowels are produced by American English speakers. The same pattern is observed in the production of male speakers as compared to that of female speakers.

Table 25: Significant and non-significant effects for CQ (Marginal $R^2 = 0.138$ /Conditional $R^2 = 0.498$).

| Backward reduced random-effect table: | | | |
|--|-------------------|----------------|----------------------|
| | <i>Eliminated</i> | <i>AIC</i> | <i>Pr(>Chisq)</i> |
| (1 Speaker) | 0 | -147.04 | 9.207e-09 *** |
| Backward reduced fixed-effect table: | | | |
| | <i>Eliminated</i> | <i>F value</i> | <i>Pr(>F)</i> |
| Language:Gender | 1 | 0.0523 | 0.8228 |
| Language | 2 | 0.7436 | 0.4052 |
| Gender | 3 | 3.0434 | 0.1062 |

The model that was found for CQ was the following:

$$CQ \sim (1 | \text{Speaker})$$

We fitted a linear mixed-effects model with CQ as dependent variable. Vowel and Vowel type

were excluded because they did not account for enough variance. Results, reported in Table 25, show that there was no effect of Language, Gender or of the interaction between Language and Gender. Interestingly, we observed a significant effect of Vowel whereas no effect was found in the model fitted with H1*-H2*. Speaker effect is represented in Figure 42. There is a lot of inter-individual variability observed in the production of FM_06, as well as in that of AM_03, although not as much. FF_01 can be considered the odd one out as she produces the lowest CQ values, indicating overall greater glottal opening (larger OQ). *Post-hoc* comparisons showed that FF_01 ($M = 0.53 \pm 0.06$) produced significant lower CQ values than FF_08 ($M = 0.83 \pm 0.06$), AM_01 ($M = 0.86 \pm 0.07$), and AM_02 ($M = 0.85 \pm 0.06$, $p < .05$).

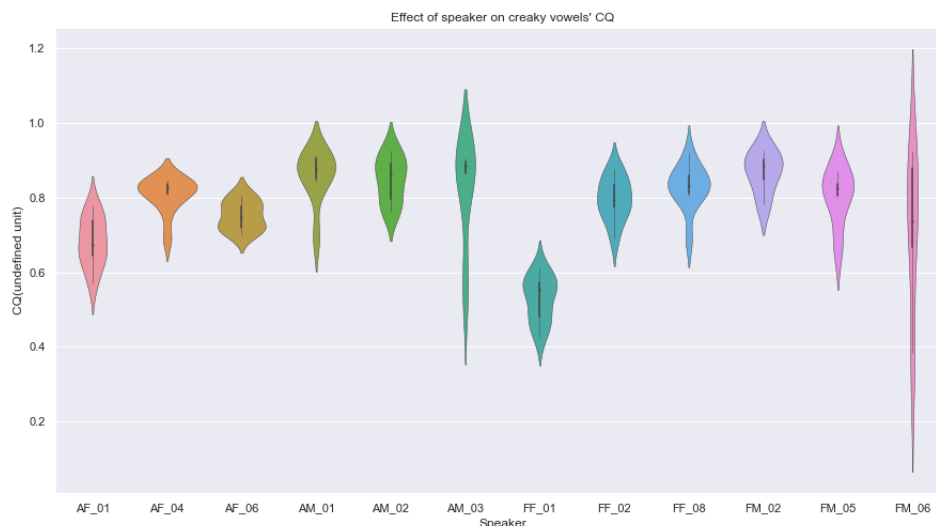


Figure 42: Violin plot representing the distribution of creaky vowels' CQ values across model speakers.

5.1 PIC

The distribution of creaky vowels' PIC values across language and gender is displayed in Figure 43. Overall, there appears to be more variability in the production of American English speakers than there is in the production of French speakers. Male American English speakers

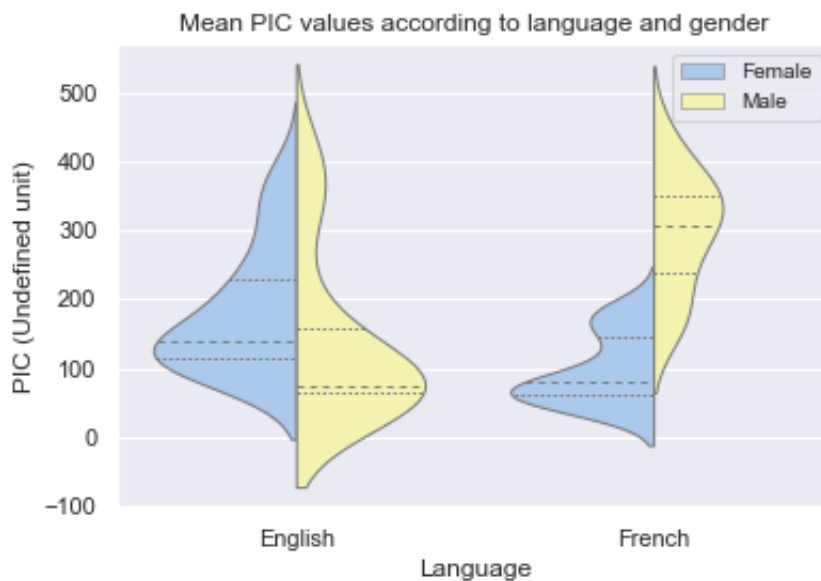


Figure 43: Violin plot representing the distribution of creaky vowels' PIC values across language and gender.

produce creaky vowels with lower PIC values, with a more prominent peak around 70, than their female counterparts for which the peak is located around 130. The distribution of values is, however, quite similar, and both male and female American English speakers seem to produce equally high values. Male French speakers seem to produce creaky vowels with higher PIC values, with a peak located around 340, than their female counterparts for which the most prominent peak is located around 70.

Table 26: Mean PIC values across language and gender. Standard deviation is indicated in parenthesis. **FM** is for French males, **FF** for French females, **AM** for American males, and **AF** for American females

| Mean PIC and SD (undefined unit) | | |
|----------------------------------|----------------|-----------------|
| Language | <i>French</i> | 185.44 (121.1) |
| | <i>English</i> | 162.92 (106.62) |
| Gender | <i>Male</i> | 213.24 (131.97) |
| | <i>Female</i> | 140.88 (84.07) |
| Language:Gender | <i>FM</i> | 299.51 (85.38) |
| | <i>FF</i> | 97.99 (50.34) |
| | <i>AM</i> | 139.75 (120.25) |
| | <i>AF</i> | 183.77 (89.65) |

We can observe in Table 26 that, overall, creaky vowels when produced by French speakers have a higher mean PIC value than when produced by American English speakers. The same observation can be made for male speakers as compared to female speakers. PIC is also higher for creaky vowels produced by male French speakers compared to when produced by their female counterparts. The opposite direction is observed in the speech of American English speakers. As reported by Esposito (2012), PIC is thought to be correlated with vocal fold closure and a lower PIC value with creakier phonation (slower vocal fold vibrations). Our observations suggest that American English and female speakers' vocal folds vibrate slower in the production of creaky vowels. Vocal fold vibration is also slower for creaky vowels produced by female French speakers as compared to their male counterparts. The reverse is observed for American English speakers.

Table 27: Significant and non-significant effects for PIC (Marginal $R^2 = 0.420$ /Conditional $R^2 = 0.716$).

| Backward reduced random-effect table: | | | |
|--|-------------------|----------------|----------------------|
| | <i>Eliminated</i> | <i>AIC</i> | <i>Pr(>Chisq)</i> |
| (1 Speaker) | 0 | 1306.0 | 3.549e-12 *** |
| Backward reduced fixed-effect table: | | | |
| | <i>Eliminated</i> | <i>F value</i> | <i>Pr(>F)</i> |
| Language:Gender | 0 | 11.334 | 0.005568 ** |

The model that was found for PIC was the following:

$$\text{PIC} \sim \text{Language} + \text{Gender} + \text{Language:Gender} + (1 \mid \text{Speaker})$$

We ran a linear mixed-effects model with PIC as dependent variable. Vowel and Vowel type were excluded because they did not account for enough variance. Results reported in Table 27 show there were no Language and Gender effects, but the interaction between both variables was found to be significant. We can observe in Figure 44 that male American English speakers produce creaky voice with lower values than their female counterparts, but that the reverse is observed in French. *Post-hoc* pairwise comparisons showed that female French speakers produced

creaky vowels with significantly lower values ($p < 0.05$) than male French speakers, therefore with significantly higher speech of vocal fold closure. No other significant difference was found.

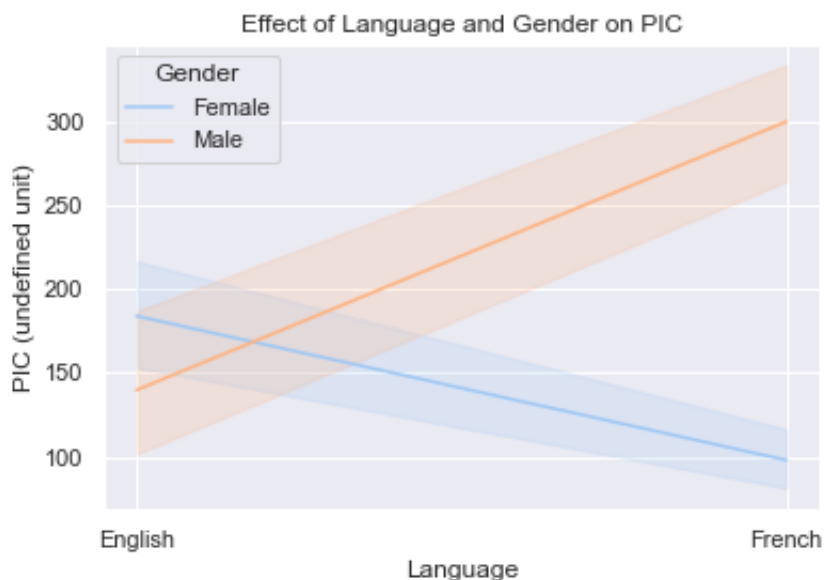


Figure 44: Interaction effect between language and gender on creaky vowels' PIC.

Speaker effect is illustrated in Figure 45. What is striking here is that there is a great deal of inter-individual variability among the majority of the speakers. Although some speakers seem to produce creaky vowels with consistent speed of vocal fold closure, this indicates that others can produce creaky voice with various speed of vocal closure speed. *Post-hoc* comparisons showed that FF_02 ($M = 56.6 \pm 10.64$) produced significantly lower PIC values than FM_02 ($M = 216.11 \pm 199.03$; $p < .05$), FM_05 ($M = 210.47 \pm 96.58$), AM_03 ($M = 221.74 \pm 133.35$; $p < .01$), AF_01 ($M = 289.36 \pm 70$), and FM_06 ($M = 305.52 \pm 111.75$; $p < .001$); AM_01 ($M = 62.98 \pm 5.4$) and FF_01 ($M = 74.58 \pm 17.44$) produced significantly lower PIC values than FM_05 ($M = 210.47 \pm 96.58$), FM_02 ($M = 216.11 \pm 199.03$), AM_03 ($M = 221.74 \pm 133.35$; $p < .05$), AF_01 ($M = 289.36 \pm 70$), and FM_06 ($M = 305.52 \pm 111.75$; $p < .001$); AM_02 ($M = 105.12 \pm 110.01$) and AF_04 ($M = 109.75 \pm 9.8$) produced significantly lower PIC values than AF_01 ($M = 289.36 \pm 70$) and FM_06 ($M = 305.52 \pm 111.75$; $p < .001$); AF_06 ($M = 152.2 \pm 36.32$) produced significantly lower PIC values than AF_01 ($M = 289.36 \pm 70$; $p < .05$)

and FM_06 ($M = 305.52 \pm 111.75$; $p < .01$); FF_08 ($M = 162.79 \pm 23.9$) produced significantly lower PIC values than FM_06 ($M = 305.52 \pm 111.75$; $p < .05$).

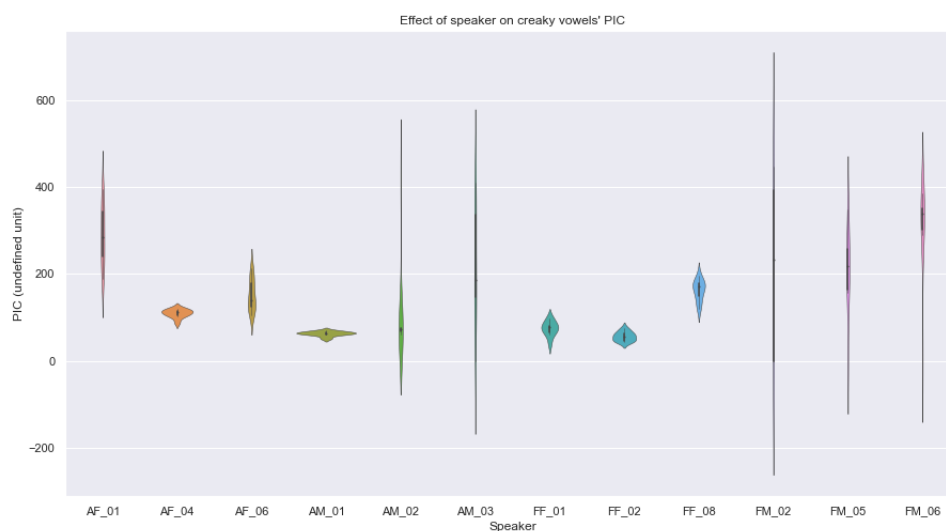


Figure 45: Violin plot representing the distribution of creaky vowels' PIC values across model speakers.

In this section we analysed the model speakers' production to observe whether cross-language variations and cross-gender differences emerged in the acoustics of creaky voice. f_0 and PIC significantly distinguished creaky voice across language and gender. $H2^*-H4^*$ and CPP distinguished creaky vowels across gender, and vowel duration across language. No other EGG or acoustic measures distinguished creaky vowels across language or gender. A lot of variability was observed in our dataset but this analysis did not provide information as to which variables accounted for the most variance, or whether those variables were shared across language. Principal component analyses were performed and are reported in the next section. Correlations between acoustic and EGG measures were also analysed. Gender is only included in the last part of the analysis.

6 Principal component analyses

Global principal component analyses were performed on the whole dataset to capture which variables were responsible for more variance across language. Cross-gender differences were then analysed within-language. Data needs to be normalised before conducting principal component analyses. We detail the method used in section 6.1. Primary observations are given in section 6.2. Correlation and variance analyses are reported in section 6.3 and 6.4, respectively. The analysis of within-language cross-gender differences is conducted in section 6.5.

6.1 Data normalisation

Data needs to be normalised to allow the comparison of quantitative variables. Normalisation was done using the Min-Max method⁵⁷. Following Lee *et al.* (2019: 1570), “[...] for each speaker, the obtained values of each acoustic variable were normalized with respect to the overall minimum and maximum values from the entire set of voice samples from males or females, as appropriate, so that all variables ranged from 0 to 1”.

6.2 Preliminary analysis

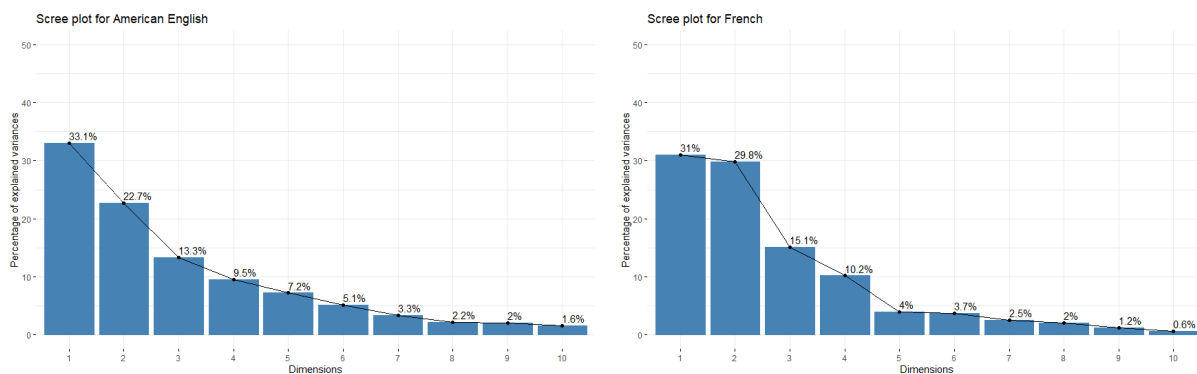


Figure 46: Scree plot representing the number of dimensions and their corresponding percentage of explained variances resulting from the PCA analyses of French and American English.

⁵⁷For each feature, the minimum value gets transformed into a 0, and the maximum value into a 1. Every other value gets transformed into a decimal between 0 and 1.

Analyses were conducted separately for French and American English and resulted in ten principal components (PCs), as seen in Figure 46. Although the Kaiser criterion states that all components with eigenvalues greater than 1 should be retained, only the first two PCs accounting for the variance of creaky voice in our data were analysed⁵⁸. They were responsible for approximately 55% of the cumulative acoustic variance in American English, and 61% in French. For best analysis, the selected PCs should be able to describe at least 75%-80% of the variance.

6.3 Correlation analysis

Correlation circles for each language were designed to visualise the distances or correlations between our different variables. They are represented in Figure 47. A high \cos^2 value indicates a good representation of the variable on the selected PCs, meaning that this variable accounts for more variability. In such cases, the variable is located nearby the outer limits of the correlation circle. Conversely, a low \cos^2 value indicates a poor representation of the variable on the selected PCs, meaning that this variable accounts for less variability. The variable is located close to the center of the correlation circle.

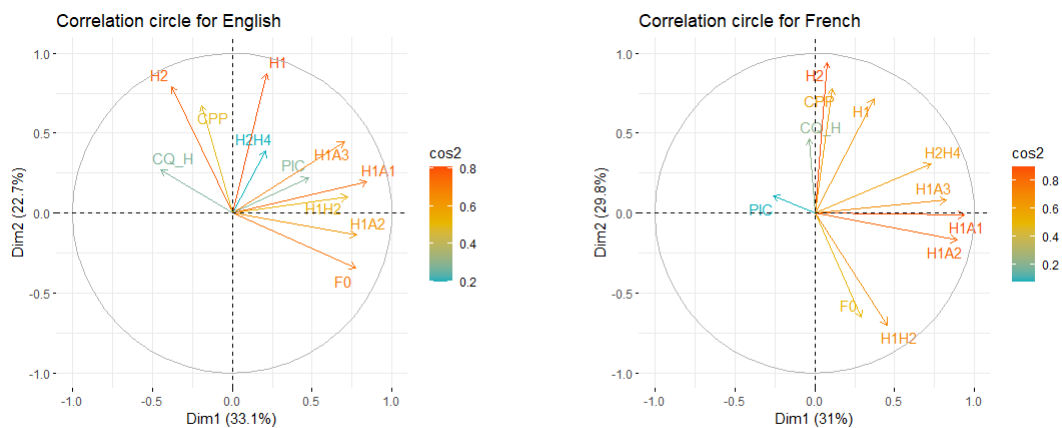


Figure 47: Correlation circles for American English and French.

We can observe that variables accounting for most variance in American English are, in a descending order, $H1^*$, $H2^*$, $H1^*-A1^*$, $H1^*-A3^*$, and f_0 . Variables accounting for least variability

⁵⁸Dimensions 3 and 4 also had an eigenvalue greater than 1 but a thorough analysis on correlations and principal components is beyond the scope of this study. Preliminary observations could, however, lead to future research.

are, in an ascending order, $H2^*$ - $H4^*$, CQ, and PIC. Variables accounting for most variability in French are, in a descending order, $H2^*$, $H1^*$ - $A1^*$, and $H1^*$ - $A2^*$, while those accounting for least variability are, in an ascending order, PIC and CQ. Variables pointing in the same direction are positively correlated, variables orthogonal to one another are unrelated, and variables pointing in opposite directions are negatively correlated. Similar sets of correlated variables can be identified across language, as displayed in Figure 47. To give an example, variables positively correlated in both languages are, among others, $H2^*$, CPP and CQ; $H1^*$ - $A1^*$, $H1^*$ - $A2^*$, and $H1^*$ - $A3^*$; f_0 and $H1^*$ - $H2^*$, and variables negatively correlated are, among others, CQ with f_0 and $H1^*$ - $H2^*$, or $H2^*$ and f_0 . Other correlations between acoustic variables could be pointed out but, following Esposito (2012), we only focused on correlations between acoustic and EGG measurements.

CQ seems to be overall correlated with the same variables across French and American English. It is positively correlated with $H1^*$, $H2^*$ and CPP, but the correlation is higher in French. CQ is also orthogonal, or almost orthogonal, with $H1^*$ - $A3^*$ and $H2^*$ - $H4^*$. Finally, it is negatively correlated with $H1^*$ - $H2^*$ and f_0 , as well as with $H1^*$ - $A1^*$ and $H1^*$ - $A2^*$, but to a lesser extent. It is not surprising to observe such correlations. When CQ increases, the vocal folds spend more time approximated than apart. Since it is negatively correlated with $H1^*$ - $H2^*$ and f_0 , when CQ increases, then $H1^*$ - $H2^*$ and f_0 decrease. The higher the closed quotient, the higher the degree of constriction and the lower the f_0 , which corroborate the acoustic description of some types of creaky voice (Keating *et al.*, 2015). If we now turn to the negative correlations, it means that the more time the vocal folds spend approximated, the smaller the posterior glottal opening will be (low $H1^*$ - $A1^*$), and the more skewed the shape of the glottal pulse will be, indicating more abrupt changes (low $H1^*$ - $A2^*$). These results are in line with observations made by Childers & Lee (1991) on vocal fry, and partially in line with those made by Esposito (2012). We both observed negative correlations between CQ and $H1^*$ - $H2^*$, $H1^*$ - $A1^*$, and $H1^*$ - $A2^*$, as well as no correlation between CQ and $H2^*$ - $H4^*$. Our results disagree on $H1^*$ and $H2^*$ for which we observed positive correlations while she observed negative and no correlation, respectively. $H1^*$ being an indirect measure of glottal aperture, we would have expected to observe a negative correlation with CQ, like she did.

Correlations for PIC, on the other hand, do not always go in the same direction across

language. It is positively correlated with $H1^*$, $H2^*$ and CPP in French, and to $H1^*$ and CPP in American English (but is orthogonal with $H2^*$). Consistent with Esposito (2012), PIC is negatively correlated with $H1^*-A1^*$, $H1^*-A2^*$ and $H1^*-A3^*$ in French, suggesting that these measures might not reflect the same aspect of speech production.

If creaky phonation does indeed have a slower vocal fold closure, then this should also affect the spectral measures $H1^*-A1^*$, $H1^*-A2^*$, $H1^*-A3^*$. These measures work under the principle that faster vocal fold closure excited the higher frequencies of a vowel, making $A1^*/A2^*/A3^*$ greater than $H1^*$. If the vocal folds are vibrating more slowly during creaky phonation [lower $H1^*-A1^*/A2^*/A3^*$], then what is causing the value of $A1^*/A2^*/A3^*$ to be greater than that of $H1^*$?

Esposito (2012: 475)

PIC is also negatively correlated with f_0 and $H2^*-H4^*$. The reverse pattern is observed in American English in which PIC is actually positively correlated with $H1^*-A1^*$, $H1^*-A2^*$, and $H1^*-A3^*$. PIC is also negatively correlated with CQ in American English, whereas the correlation is positive in French, suggesting there is a relationship between the degree of vocal fold opening and PIC, but that this relationship is inverted across language. Esposito (2012: 475) also found a negative correlation between PIC and CQ which she explained by the fact that “[c]reaky phonation [...] is produced by vocal folds that are close together, and therefore do not need to move as quickly to reach a state of closure. [...] The more contact between the vocal folds, the higher the CQ value, but the lower the [PIC] value”. However, CQ and PIC accounted for less variability in our dataset, which means that other dimensions than the degree of glottal opening and the speed of glottal closure are responsible for more variability in the production of creaky voice.

6.4 Variance analysis

The contribution of variables to PC1 and PC2 in both languages is displayed in Figure 48. Variables below the red line do not quite explain variance in a given PC and can, for this reason, be eliminated. We can observe that, in PC1, the 5 variables $H1^*-A1^*$, $H1^*-A2^*$, f_0 , $H1^*-H2^*$ and $H1^*-A3^*$, account for most variance in American English. There are only 4 variables in French

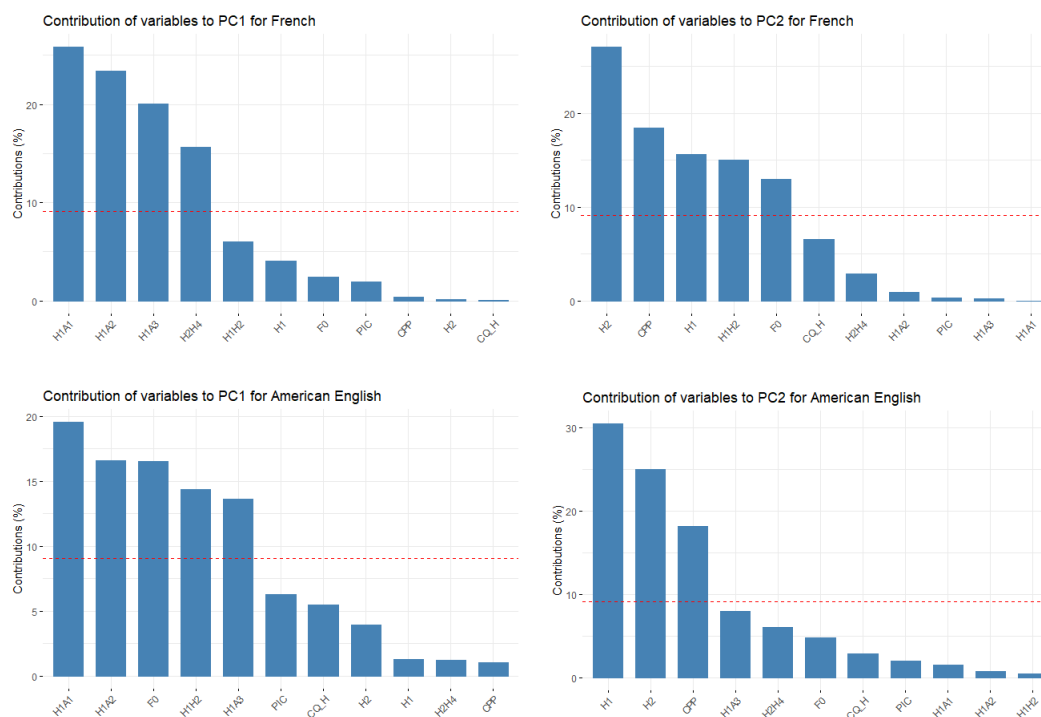


Figure 48: Contribution of variables for PC1 and PC2 in both French and American English.

which are H1*-A1*, H1*-A2*, H1*-A3*, and H2*-H4*. In PC2, H1*, H2*, and CPP account for most variance in American English and H2*, CPP, H1*, H1*-H2*, and f_0 in French. The general picture that emerges from these observations is a somewhat similar acoustic organisation of creaky voice across language for the first two PCs were largely shared. H1*-A1*, H1*-A2*, and H1*-A3* for instance, accounted for most variability in both languages in PC1, which implies that creaky voice might be realised with variable speed of vocal fold closure. Between-language differences were also observed. f_0 accounts for most variability in PC1 in American English and in PC2 in French, suggesting that f_0 varies more in American English creaky voice than in French creaky voice. The same observation can be made for H1*-H2*. Considering there are sub-types of creaky voice that do not have a low f_0 (e.g. tense voice), or that are non-constricted (cf. Slifka voice), this could indicate that there might be less sub-types of creaky voice, or less ways of realising creaky voice in French than there are in American English. Conversely, H2*-H4* accounts for most variability in PC1 in French but does not in either analysed PCs in American English. H1*, H2* and CPP all account for most variability in PC2 across language. Disparity in

vowel quality across our dataset could explain the observed variance in $H1^*$ and $H2^*$. However, the fact that they are positively correlated (to some extent) shows that the difference between $H1^*$ and $H2^*$ remains quite low, hence does the degree of constriction. Variations in CPP might bring further evidence that different types of creaky voice are present in our dataset. Indeed, CPP is a measure of periodicity, and not all types of creaky voice have an aperiodic signal (e.g. vocal fry).

6.5 Cross-gender differences within-language analysis

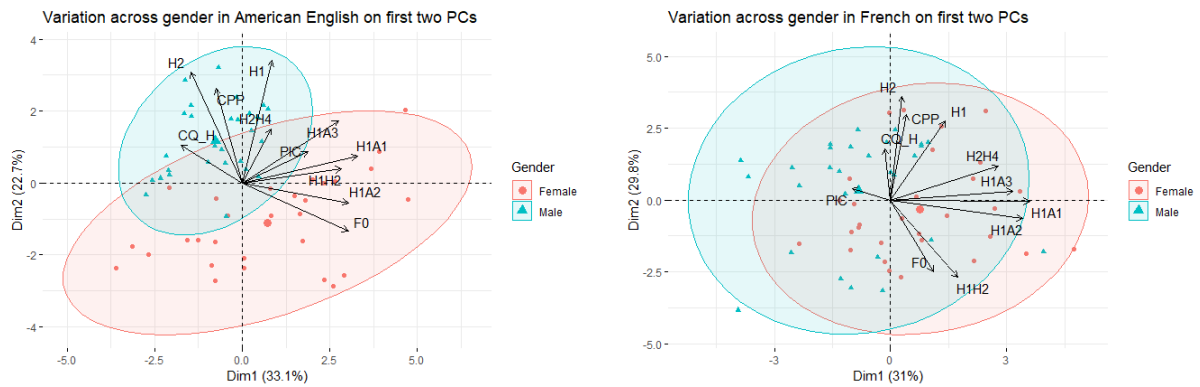


Figure 49: Variation across language and gender in both PC1 and PC2.

Our analysis of principal components have only taken into account cross-language differences so far. However, gender might play a substantial role in the variance observed across language. This can be shown by plotting individual factor maps around the qualitative variable Gender. The confidence ellipsis, as can be observed in Figure 49, of the two modalities (male/female) seem to suggest that these two categories are more differentiated in English and overlapping in French. There is much more variation in the production of female speakers than in that of male speakers in American English, meaning that there is more variability in the way female speakers produce creaky voice. Conversely, there is far less variability in French and the ellipses do not differ much in either size or in the way they superimpose. Spectral measures correlated with vocal fold closure, $H1^*$ - $H2^*$, as well as f_0 , seem to account for most variability in the production

of female American English speakers, while more variability is observed in CPP, H1* and H2* in male American English speakers production. This could indicate that female American English speakers produced more sub-types of creaky voice as compared to their male counterparts, and that creaky voice is less gender differentiated in French.

Conclusions drawn from our PCA analyses confirm that there exist cross-language differences in the realisation of creaky voice, and that some cross-gender acoustic differences might be language-dependent. We observed substantial variation in the production of female American speakers as compared to that of their male counterparts. Gender-based differences in French were not significant in comparison. Our observations support the idea that physiology alone is unlikely to explain acoustic differences observed in creaky voice, and that such behaviours must be socially constructed. The social construction of creaky voice also seems to be more predominant in American English, and/or less sub-types of creaky voice might exist in French.

7 Conclusion

In this chapter we provided a descriptive and comparative analysis of the acoustic and physiology of French and American English creaky voice. Cross-gender analyses were conducted within-language. We included Speaker, Vowel and Vowel type in our statistical analyses. Not all variables accounted for the same effect depending on measurement. A synthesis of the models found for each variable and their significant effect is given in Table 28. CPP and f_0 significantly distinguished creaky vowels between languages, f_0 and H2*-H4* significantly distinguished creaky vowels across gender, and significant cross-gender differences were only observed for PIC in French. A lot of intra- and inter-variability was observed across the entire dataset, which corroborates the idea that there is a myriad of subtle configurations that can be achieved to produce creaky voice, and phonation types in general. Both EGG measures displayed very little variance in both French and American English. Other acoustic dimensions, like H1*-A1*, H1*-A2*, and H1*-A3*, revealed a lot of variability in both languages, indicating that creaky voice might be realised with variable speed of vocal fold closure. f_0 and H1*-H2* were found to vary more in American English creaky voice than in French creaky voice, which could indicate that Ameri-

can English speakers produced more sub-types of creaky voice, or that not as many sub-types exist in French creaky voice. CQ was correlated with the same variables in both languages but correlations for PIC went in the opposite direction. Cross-gender differences in the production of creaky voice are also much more apparent in American English: there was substantially more variability in the way female speakers produce creaky voice than in the way male speakers do. No striking cross-gender differences were observed in French apart from PIC. There might be more sub-types of creaky of creaky voice existing in American English, with female producing more of them. These cross-gender differences also confirm that creaky voice usage is a socially constructed phenomenon, and that it is more predominant in American English than in French.

Vowel and Vowel type effects should be carefully interpreted for the number of tokens in each category is uneven. More research using better controlled data is needed to further explore the relationship between vowel quality and voice quality.

All of our conclusions are based on a very small sample size and on averaged measurements made over the course of the entire vowel. Measurements taken at different points might result in more accurate observations seeing that non-modal phonation types are often localised to a portion of the vowel only. Therefore, our results might not be representative of general tendencies observed in previous studies.

In the next chapter we analyse creaky voice accommodation and evaluation. The analysis is divided in sections relative to the three tasks we conducted.

Table 28: Synthesis of the models found for each variable and their significant effects.

| Measures | Models found | Effects | |
|----------|--------------|--|--|
| ACOUSTIC | $\log(dur)$ | $dur(\log) \sim \text{Language} + \text{Gender} + \text{Language:Gender} + (1 \text{Speaker}) + (1 \text{Vowel_type}) + (1 \text{Vowel})$ | Vowel_type; $p < .05$ Language:Gender; $p < .05$ Speaker; $p < .001$ |
| | f_0 | $F0 \sim \text{Language} + \text{Gender}$ | Language; $p < .05$ Gender; $p < .001$ |
| | $H1^*$ | $H1 \sim (1 \text{Speaker}) + (1 \text{Vowel})$ | Speaker; $p < .001$ Vowel; $p < .001$ |
| | $H2^*$ | $H2 \sim (1 \text{Speaker}) + (1 \text{Vowel_type})$ | Speaker; $p < .001$ Vowel_type; $p < .001$ |
| | $H1^*-H2^*$ | $H1H2 \sim 1$ | None |
| | $H1^*-A1^*$ | $H1A1 \sim (1 \text{Speaker}) + (1 \text{Vowel})$ | Speaker; $p < .001$ Vowel; $p < .001$ |
| | $H1^*-A2^*$ | $H1A2 \sim (1 \text{Speaker}) + (1 \text{Vowel})$ | Speaker; $p < .01$ Vowel; $p < .001$ |
| | $H1^*-A3^*$ | $H1A3 \sim (1 \text{Speaker}) + (1 \text{Vowel})$ | Speaker; $p < .001$ Vowel; $p < .001$ |
| | $H2^*-H4^*$ | $H2H4 \sim \text{Gender} + (1 \text{Vowel_type})$ | Gender; $p < .001$ |
| | CPP | $CPP \sim \text{Language} + (1 \text{Speaker}) + (1 \text{Vowel_type})$ | Language; $p < .05$ Vowel_type; $p < .01$ Speaker; $p < .001$ |
| EGG | CQ | $CQ \sim (1 \text{Speaker})$ | Speaker; $p < .001$ |
| | PIC | $PIC \sim \text{Language} + \text{Gender} + \text{Language:Gender} + (1 \text{Speaker})$ | Language:Gender; $p < .01$ Speaker; $p < .001$ |

CHAPTER V: Accommodation and perception of creaky voice

Phonetic accommodation and voice quality have been extensively studied these past decades, but there are still very few studies on accommodation of voice quality, and even less conducted cross-linguistically. To our knowledge, this study is the first multidimensional analysis on creaky voice accommodation among French learners of English.

To determine whether female French learners of English successfully accommodated to creaky voice, we first analyse their creaky voice usage in both French and English in a reading task. We then compared this data to that obtained from the repetition/imitation task to observe any convergence effects. All twelve acoustic and EGG measures were compared. We included Language and Gender in our analysis to observe any possible between-language and within-language cross-gender differences in accommodation patterns. A rating task was conducted to determine whether creaky voice was more positively evaluated than non-creaky voice by female French learners of English, and whether Language and Gender significantly influenced judgments of creaky voice. Voice quality was rated on four different traits: pleasant, attractive, powerful, and educated.

CHAPTER CONTENT

This chapter is organised around our three main tasks. In the first section we compare creaky voice usage across language in the reading task. We first report general observations and then split the data to observe between-speaker variability and Vowel type effect. In the third section we analyse convergence effects observed in the repetition/imitation task. We first describe how we measured convergence and then report our results for every acoustic and electroglottographic measurements we made. Analyses of global convergence and of convergence across language and gender are provided. The fourth section deals with the evaluation of creaky vs. non-creaky voice. As for the repetition/imitation task, we first report global ratings of creaky vs. non-creaky voice. We then included between-language and cross-gender analyses to observe any differences in how creaky voice might have been perceived across these different groups. We also included a section for Model speaker effect to show that listeners relate to other perceptual features than just voice quality when judging a voice.

1 Reading task

Before addressing the question of creaky voice accommodation, we analysed whether vowels contained in word-final position were more subject to being produced in a creaky or non-creaky voice when subjects read in French or in English. We extracted all vowels of interest and calculated the percentage of vowels that were creaky as compared to those that were not. 17 sentences were read by 20 subjects in each language. 340 sentences were, therefore, read in each language. 679 and not 680 vowels were extracted, for S10 failed to read one sentence. Overall observations are provided in section 1.1. We then analysed between-speaker variability in section 1.2 and Vowel type effect in section 1.3.

1.1 Overall observations

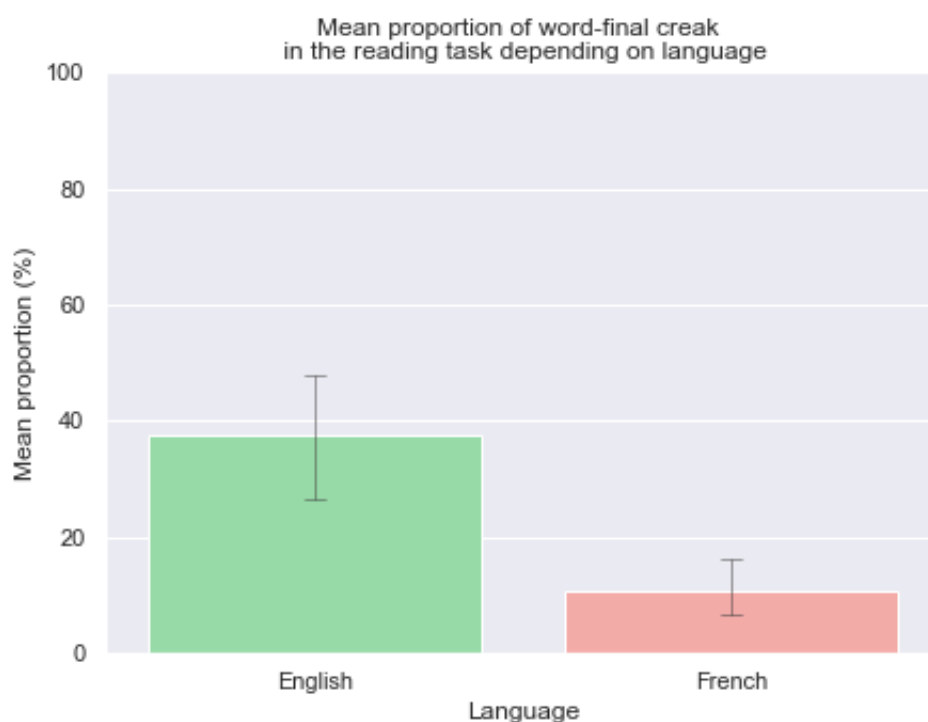


Figure 50: Mean proportion of creaky vowels in the reading task depending on language. Error bars represent the standard error of the mean.

Figure 50 shows the mean proportion of word-final creak across language. We can observe that the proportion of word-final creak is higher when subjects read in English than when they read in French. A Wilcoxon paired t-test resulted in significant effect of Language on the proportion of word-final creak ($z_{20} = 3.25$; $p < .01$). Subjects produced significantly more occurrences of word-final creak in English ($M = 39.01 \pm 25.21$) than they did in French ($M = 10.86 \pm 11.81$). We can conclude that French learners use more instances in creak when reading in English than when reading in French. Although our analysis focuses only on one specific locus, this corroborates observations made by Pillot-Loiseau *et al.* (2019) who observed a higher proportion of creaky occurrences when subjects read in L2 English than when reading in L1 French.

1.2 Between-speaker variability

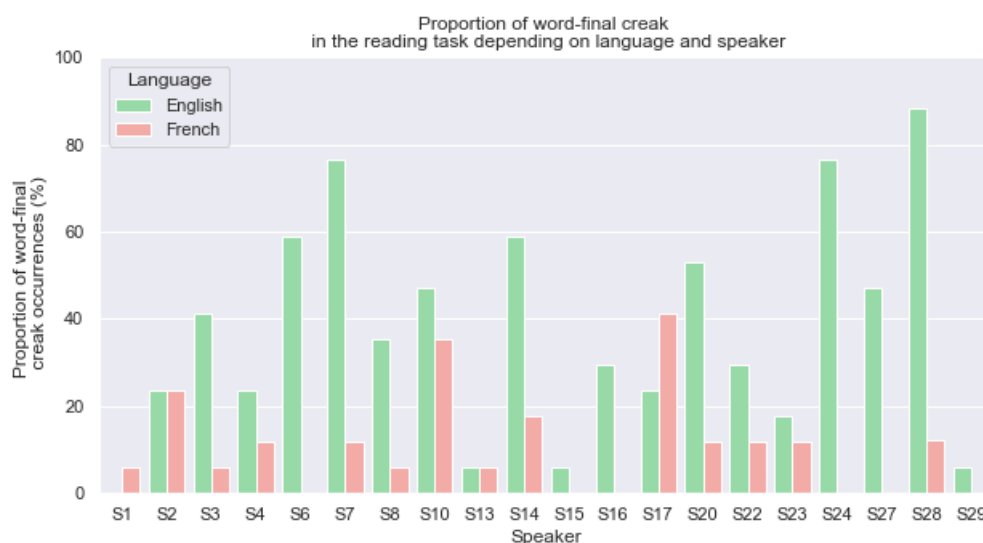


Figure 51: Mean proportion of word-final creak in the reading task depending on language and speaker.

As always in speech, inter-speaker variability is inevitable. The general tendency previously observed might not be reflected in each subject. We can observe in Figure 51 that inter-individual variability is substantial. There are no two subjects producing the exact same proportion of word-final creak in both languages. Out of 20 subjects, 16 produced more occurrences of word-final creak when reading in English than when reading in French. Among these 16 subjects, 6 did

not produce any occurrences of word-final creak when reading in French. 2 out of 20 subjects (S13 and S2) produced the same amount of occurrences of word-final creak across language. The remaining 2 subjects (S17 and S1) produced more occurrences of word-final creak in French than in English. 1 out of them (S1) produced no occurrence of word-final creak when reading in English.

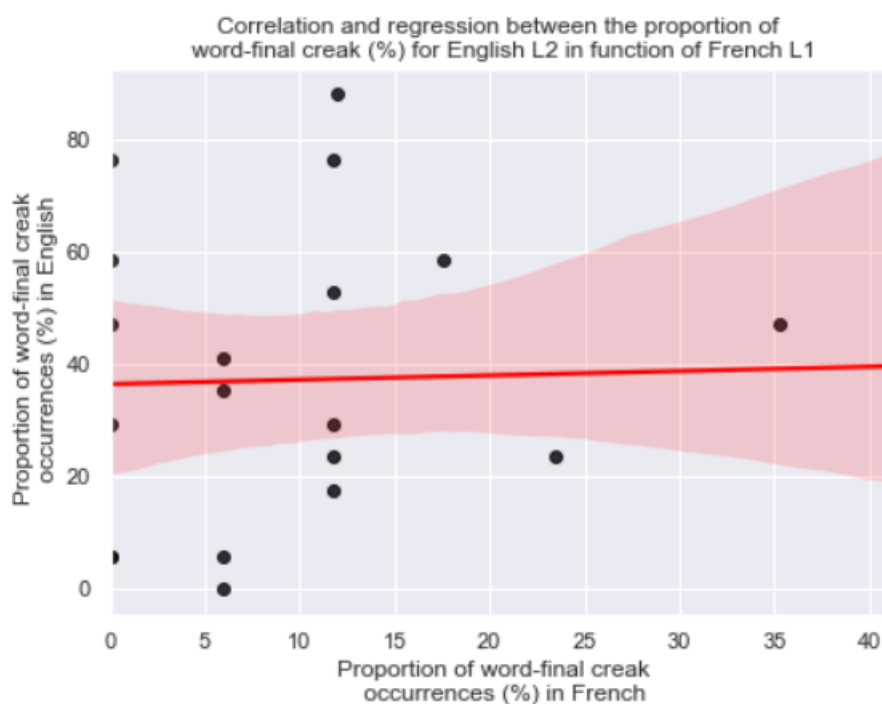


Figure 52: Correlation and regression plot showing the proportion of word-final creaky vowels (%) in English L2 as compared to that in French L1.

Following Pillot-Loiseau *et al.* (2019), we analysed whether the proportion of word-final creak in L2 English was correlated with the proportion of word-final creak in L1 French in our dataset. As can be observed in Figure 52, no significant correlation was observed ($r = 0.007$, $p > .05$), meaning that one subject producing a high number of word-final creak in French will not necessarily produce more instances of word-final creak in English.

1.3 The effect of vowel type

Due to inter-individual variability, which was again illustrated in the previous subsection, we failed to obtain an even number of words containing either a low or a high vowel for this exact reason. However, we were still interested in looking at whether vowel type had an effect on the proportion of word-final creak in our dataset. Creaky and non-creaky realisations of vowels in word-final position according to vowel type are represented in Figure 53. We can observe that, although the number of high vowels is almost double as compared to that of low vowels ($N_{High} = 440$; $N_{Low} = 239$) across the dataset, the proportion of word-final creak does not differ greatly depending on vowel type ($N_{Creaky/High} = 99$; $N_{Creaky/Low} = 74$), indicating that low vowels are probably more frequently produced with a creaky voice than high vowels. This corroborates observations made by Panfili (2015) that, due to physiological constraints, creaky voice is harder to achieve on high vowels, which is why low vowels are more often produced in a creaky voice than high vowels. Creaky high vowels might be produced with a subtype of creaky voice that involves high adductive tension (e.g. vocal fry, tense voice, etc.)

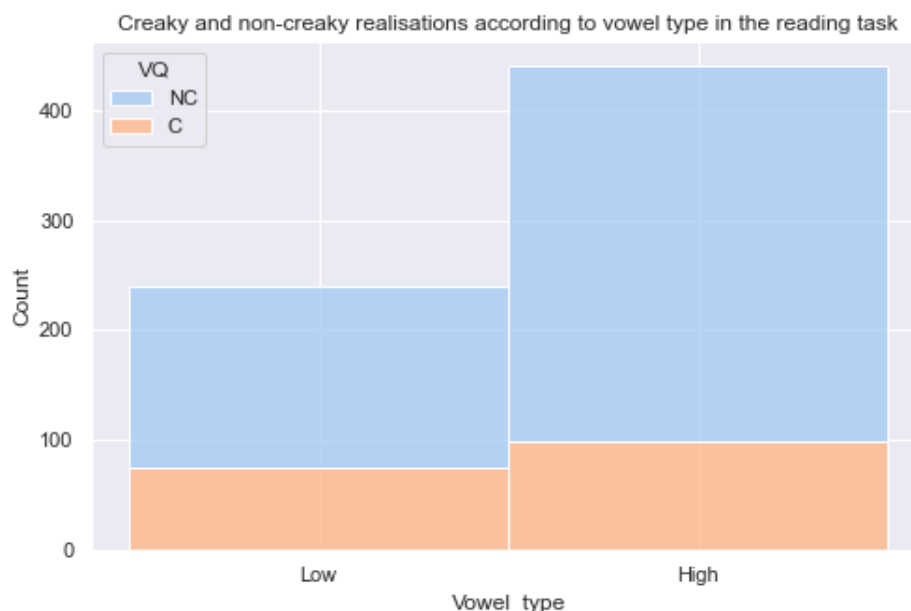


Figure 53: Creaky and non-creaky realisations of vowels according to vowel type in the reading task.

Subjects produced more instances of word-final creak when reading in English, and low vowels seemed to be more subject to creak than high vowels. We now want to observe the influence of vowel category on word-final creak across language. We have calculated the proportion of creaky low and high vowels produced by each speaker in both languages. Results are displayed in Figure 54. As previously observed, the proportion of word-final creak is overall higher in English. Wilcoxon paired t-tests resulted in significant effect of Language on the proportion of both creaky low vowels ($z_{20} = 3.18$; $p < .01$) and creaky high vowels ($z_{20} = 3.1$; $p < .01$) in word-final position. The proportion of creaky low vowels in word-final position is higher when subjects read in English ($M = 65 \pm 39.71$) than when they read in French ($M = 19.99 \pm 24.61$). The same conclusion can be drawn for high vowels ($M = 21.69 \pm 17.16$, for English; $M = 5 \pm 10$, for French). Another observation that can be made is that the number of creaky low vowels in word-final position is higher than the number of creaky high vowels in word-final position in both languages. Wilcoxon paired t-tests resulted in significant effect of vowel type on the proportion of word-final creak in English ($z_{20} = 3.38$; $p < .01$) and in French ($z_{20} = 3.31$; $p < .01$). The proportion of low vowels that are produced in a creaky voice is significantly higher than that of high vowels in word-final position, in both French and English.

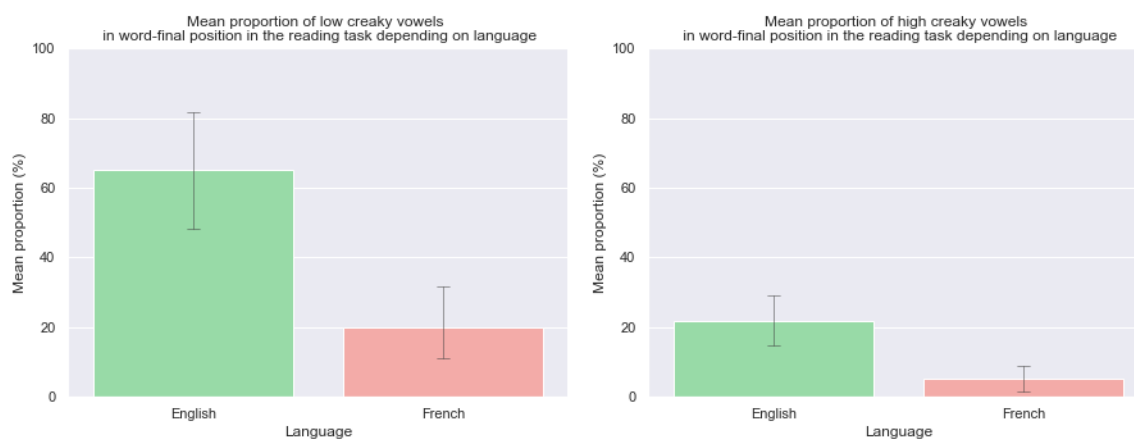


Figure 54: Mean proportion of low creaky vowels, on the left panel, and of high creaky vowels, on the right panel, across language. Error bars represent the standard error of the mean.

We also looked at how creaky and non-creaky vowels in word-final position were distributed

depending on vowels present in our dataset. We can see in Figure 55 that, even among the low/high categories, the number of occurrences per vowel varies greatly. It is clear that low vowels are more often produced in a creaky voice than high vowels, and more in English than in French. It is less apparent whether some vowels might be more subject to creak or not, but it might be the case that high back vowels in English (i.e. GOOSE) are less often creaked than high front vowels (i.e. KIT), for instance.

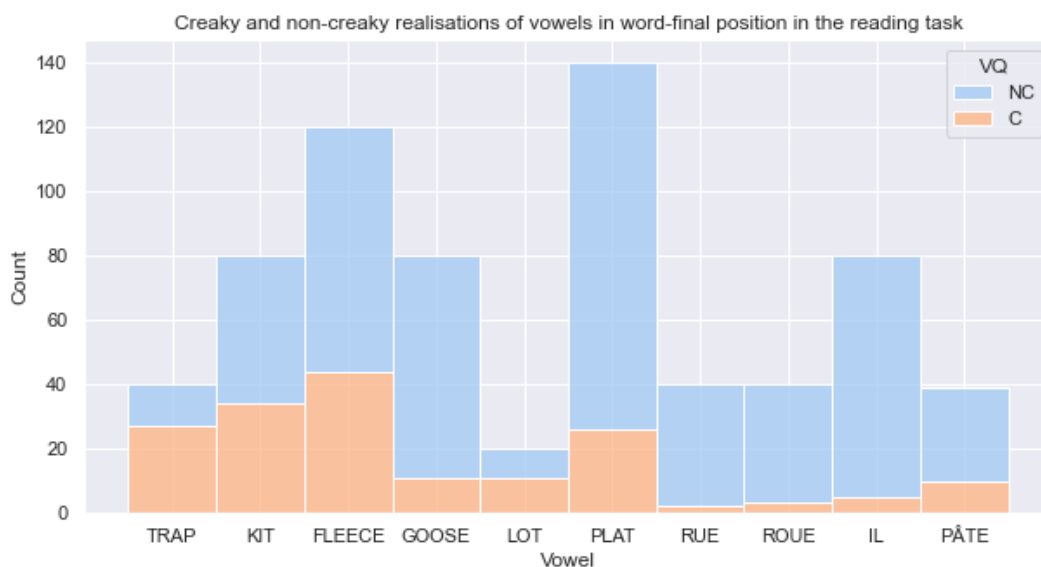


Figure 55: Creaky and non-creaky realisations of vowels in the reading task.

In this section we saw that the presence of creaky voice in read speech confirms previous observations that creaky voice usage does depend on language status (Pillot-Loiseau *et al.*, 2019). Our analysis also seems to corroborate the fact that low vowels are more often creaked than high vowels, in both languages.

In the next section we analyse creaky voice accommodation. We conducted a multidimensional acoustic and electroglottographic analysis that provided objective measurements of convergence (or divergence) effects. We looked at whether patterns of accommodation differed across language and if any cross-gender differences were observed within-language.

2 Imitation task

After conducting the reading task, subjects conducted a repetition task in which they were explicitly instructed to imitate what they heard. They had no further indication as to what speech features they should imitate. They heard twice the same sentences produced by each model speaker: one containing word-final creak, the other not containing word-final creak⁵⁹. Accommodation of creaky voice being the focus of this study, we only extracted repetitions of sentences produced by model speakers that contained word-final creak. The other sentences served as distractors. We examined convergence on different acoustic and EGG dimensions to determine whether some displayed greater accommodation effects than others. Between-language differences and within-language gender-based differences were analysed. We deliberately excluded Vowel and Vowel type from this analysis due to an important disparity in each category.

2.1 Difference-in-Distance

To assess convergence effects, we used a measurement often used in accommodation studies called *difference-in-distance* (Babel, 2012; Pardo *et al.*, 2013, 2017; Lewandowski& Nygaard, 2018, Wagner *et al.*, 2021). It was calculated for all acoustic and EGG measures by subtracting the absolute shadowed difference (shadowed-model) from the absolute baseline difference (baseline-model) for each token:

$$\text{DID} = |(\text{baseline distance}) - |(\text{shadowed distance})|$$

For example, let's take the vowel /ae/ in the word *gag* that model speaker AM_03 produced with a relative f_0 of 71.756 Hz. If participant S17 produced the same vowel in that same word with a relative f_0 of 127.636 during baseline, and of 112.793 when imitating that model, their DID score for that specific token would be $|127.636 - 71.756| - |112.793 - 71.756| = 55.88 - 41.037 = 14.843$. A positive value indicates convergence (the shadowed difference is smaller than the baseline distance), a negative value indicates divergence (the shadowed difference is bigger than the baseline distance), and a 0 value indicates maintenance (no change from baseline to imitation). The magnitude of the DID scores reflects the magnitude of the effect.

⁵⁹The reader should refer to Chapter IV, section 3, for stimuli description.

2.2 Data visualization

For each measure, a bar plot illustrates the mean DID score across language and gender. Alignment towards American English model speakers' is represented on the left side and alignment towards French model speakers' on the right side. Alignment towards female model speakers is represented in blue while it is in yellow for male model speakers. Error bars represent the standard deviation of the mean. Other bar plots were created to represent the effect of Model speaker, if any. Line plots were created to represent the effect of either Language, or of the interaction between Language and Gender. For all tables, FM refers to French males, FF to French females, AM to American males and AF to American females.

2.3 Model selection

Following Lewandowski & Nygaard (2018), separate mixed-effects models (MEMs) were built to assess convergence on each acoustic and EGG measures. For each series of MEMs, we first fitted a control MEM that included only a random effect for Subject to observe overall convergence effects (`variable ~ (1 | Subject)`). Language, as well as the interaction between Language and Gender, were the fixed effects of interest in all analyses. We did not analyse Gender individually for it cannot be separated from Language⁶⁰. Model speaker was included as a random effect. Significance was determined within and between mixed-effects models with model comparison (χ^2). Significance tests used Satterthwaite's approximation for the *dfs*. We included a random intercept for Speaker and Model speaker to account for speaker- and model-dependent effects on convergence.

2.4 Analysis of acoustic measures

2.4.1 Duration DID score

Mean duration DID scores across language and gender are displayed in Figure 56. Subjects seem to have generally aligned on duration towards model speakers. Whether more convergence is observed across language is not very clear because of gender effect. If we compare duration DID

⁶⁰All subjects were female and have repeated both male and female speakers in each language. Measurements were compared between the reading and repetition tasks.

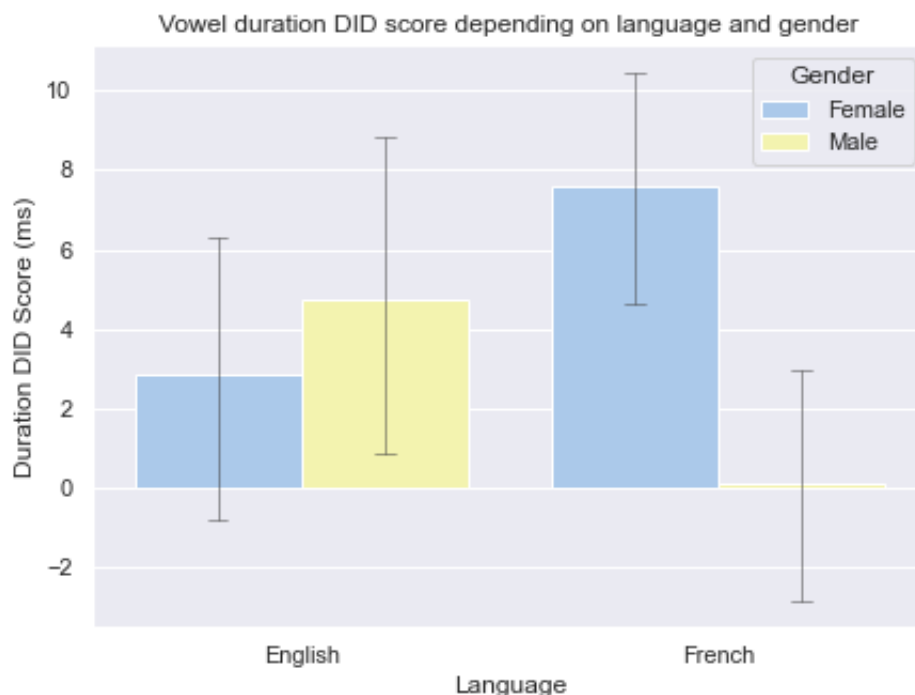


Figure 56: Bar plot representing the mean duration DID score by language and gender. Error bars represent the standard error of the mean.

score within-language, more convergence is observed towards female French speakers whereas there is almost no convergence effects towards male French speakers. The reverse pattern is observed for American English model talkers: more convergence is observed towards male than towards female models. Error bars show that divergence effects were also observed for some speakers towards female American English and male French model talkers.

Table 29 shows the mean duration difference values between the reading task and the repetition tasks according to language and model speakers. If we compare these values to the DID values displayed in Figure 56, we can conclude that subjects converged towards French and English model speakers by increasing vowel length. The same pattern is observed for French female model speakers, as well as for both male and female American English model speakers. Subjects overall decreased vowel length when imitating French male models, leading to barely any effects. It is not very clear whether subjects should have increased or decreased vowel length for convergence effects to be observed towards French male model speakers. We compared mean duration

values from the reading and repetition tasks to that obtained for French male model speakers. We concluded that subjects should have decreased vowel length even more for convergence effects to be observed.

Table 29: Mean vowel duration difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively).

| Mean duration difference (ms) | | |
|-------------------------------|----------------|------------------|
| Language | <i>French</i> | \nearrow 3.22 |
| | <i>English</i> | \nearrow 7.15 |
| Language:Gender | <i>FM</i> | \searrow 9.76 |
| | <i>FF</i> | \nearrow 16.2 |
| | <i>AM</i> | \nearrow 13.24 |
| | <i>AF</i> | \nearrow 1.4 |

To determine if Language and the interaction between Language and Gender had a significant effect on duration DID scores, we first built a control model (as described previously) with duration DID as the dependent measure. The control MEM's intercept was significant ($\beta = 3.81$, $t = 2.62$, $p < .05$), indicating significant overall convergence to duration. Language ($\beta = -0.03$, $t = -0.02$, $p = .98$; $\chi^2(1) = 2e.04$, $p = 0.98$) did not improve model fit but the interaction between Language and Gender ($\beta = -9.38$, $t = -2.72$, $p < .01$; $\chi^2(5) = 9.84$, $p < 0.05$) and Model speaker did ($\chi^2(11) = 98.11$, $p < 0.001$). *Post-hoc* comparisons showed significantly more convergence effects towards female French speakers than towards male French ones ($\beta = -9.38$, $t = 3.04$, $p < .05$). The effect is illustrated in Figure 57.

Duration alignment also significantly differed across model speakers. As can be observed in Figure 59, FF_02 ($M = 19.97 \pm 39.09$) received the most alignment on duration, followed by FM_05 ($M = 15.89 \pm 30.83$), AF_01 ($M = 15.35 \pm 40.1$), AM_01 ($M = 15.00 \pm 48.57$), AM_02 ($M = 8.51 \pm 41.38$), FF_08 ($M = 3.45 \pm 25.9$), and FF_01 ($M = 1.16 \pm 31.74$). FM_06 ($M = -11.28 \pm 38.05$) received the most divergence on duration, followed by AM_03 ($M = -7.06 \pm 48.47$), FM_02 ($M = -5.58 \pm 30.35$), AF_04 ($M = -4.00 \pm 47.06$), and AF_06 ($M = -2.76 \pm 48.09$).

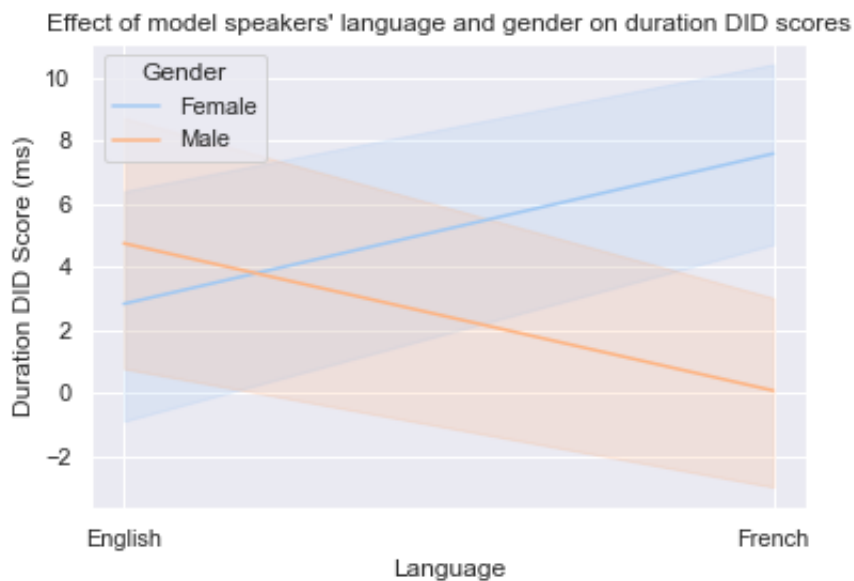


Figure 57: Effect of language and gender on duration DID scores. The shaded areas represent the standard deviation.

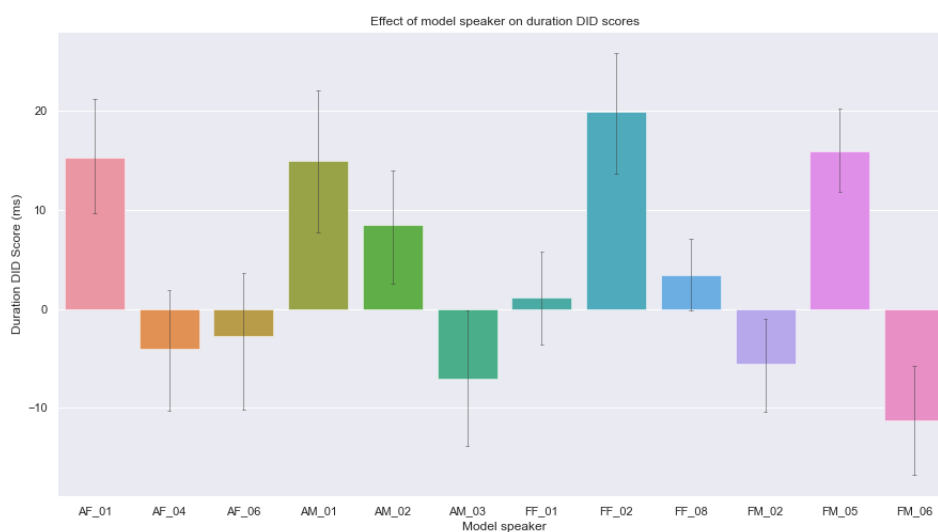


Figure 58: Effect of model speaker on duration DID scores. Error bars represent the standard error of the mean.

These results suggest that there was significant overall alignment on vowel duration. Duration alignment did not vary by language. Subjects converged towards all groups except towards French males for which overall maintenance was observed. Significant gender-based differences

were only observed in French, with subjects converging more towards female than towards male model speakers. French female model speakers received the most alignment on duration. Model speaker accounted for most variation with FF_02 receiving the most alignment on duration and FM_06 receiving the most divergence.

2.4.2 Fundamental frequency DID score

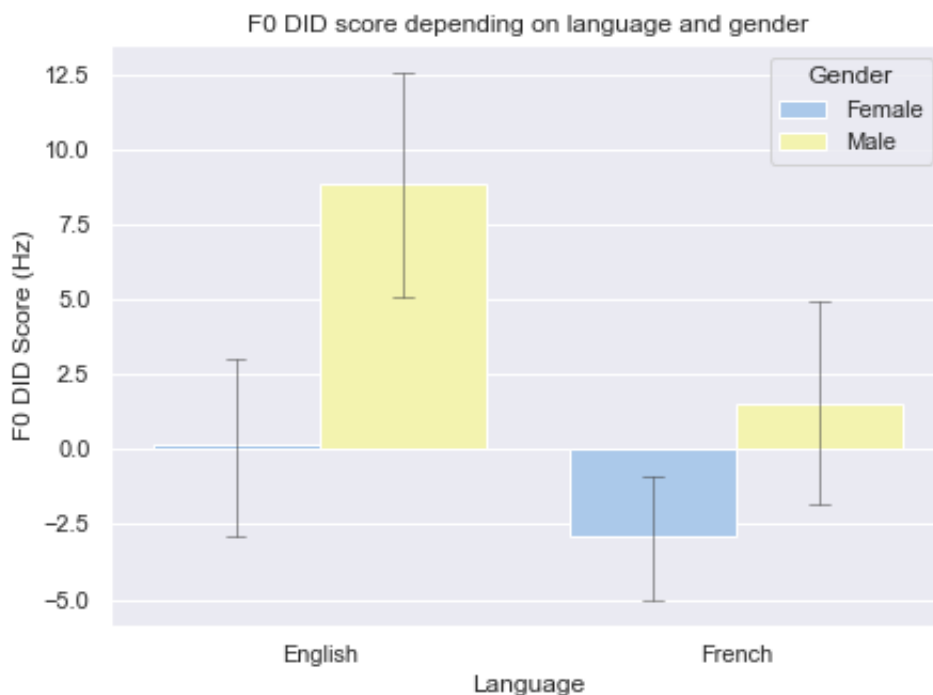


Figure 59: Bar plot representing the mean f_0 DID score by language and gender. Error bars represent the standard error of the mean.

Mean f_0 DID scores across language and gender are displayed in Figure 59. More convergence effects are observed towards American English model speakers than towards French ones. Subjects have overall converged more towards male speakers in terms of f_0 , with a stronger effect for American English male learners than towards their French counterparts. Overall maintenance is observed towards American English female speakers, while we can observe divergence towards French female speakers. Error bars show that divergence effects were also observed for some speakers towards female American English and male French model speakers.

Table 30: Mean f_0 difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively).

| Mean f_0 difference (Hz) | | |
|----------------------------|----------------|-----------------|
| Language | <i>French</i> | \searrow 3.09 |
| | <i>English</i> | \searrow 1.69 |
| Language:Gender | <i>FM</i> | \searrow 7.39 |
| | <i>FF</i> | \nearrow 1.21 |
| | <i>AM</i> | \searrow 6.66 |
| | <i>AF</i> | \nearrow 3.0 |

Table 30 shows the mean f_0 difference values between the reading task and the repetition tasks according to language and model speakers. If we compare these values to the f_0 DID values displayed in Figure 59, we can conclude that subjects converged towards both French and American English male speakers by decreasing f_0 . Subjects seem to have increased f_0 when repeating French and American English female speakers, leading to divergence effects towards French female speakers and barely any effects towards American English female model speakers. For convergence to be observed, they should have produced decreased f_0 values, at least when repeating French model speakers. Since it was not very clear whether they should have produced decreased or increased f_0 values for convergence to be observed towards female American English model speakers, we compared mean f_0 values from the reading and repetition tasks to that obtained for American English model speakers. We concluded that subjects should have produced even lower f_0 values for convergence effects to be observed towards this group.

To determine whether Language and the interaction between Language and Gender had a significant effect on f_0 , we fitted our control model (as described previously) with f_0 DID as the dependent measure. The control MEM's intercept was non-significant ($\beta = 1.94$, $t = 1.17$, $p = .26$), indicating no significant overall convergence to f_0 . Language ($\beta = -5.09$, $t = -3.17$, $p < .01$; $\chi^2(1) = 9.99$, $p < .01$), the interaction between Language and Gender ($\chi^2(5) = 28.72$, $p < .001$), and Model speaker ($\chi^2(1) = 11.35$, $p < .001$) improved model fit.

As can be observed in Figures 60 and 61, respectively, subjects converged more towards American English model speakers than towards French ones. More convergence on f_0 was also

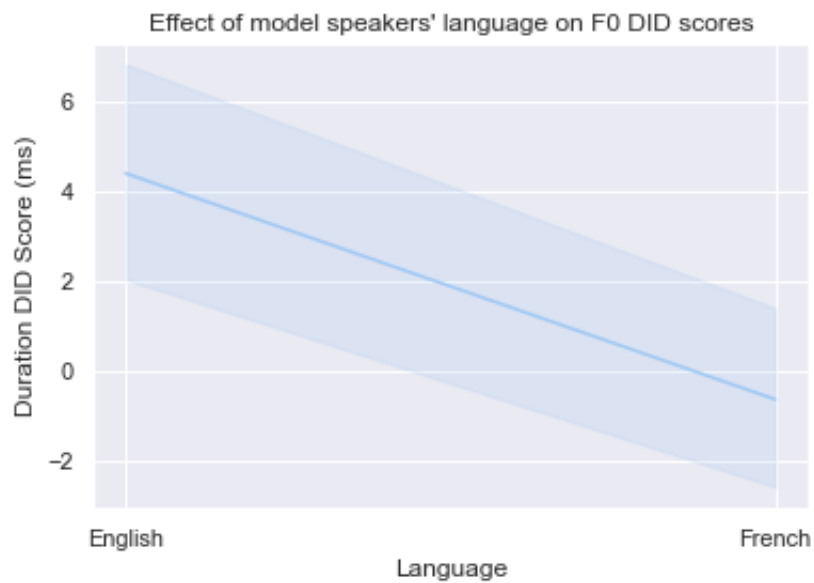


Figure 60: Effect of language on f_0 DID scores. The shaded area represents the standard deviation of the mean.

observed towards male speakers within-language.

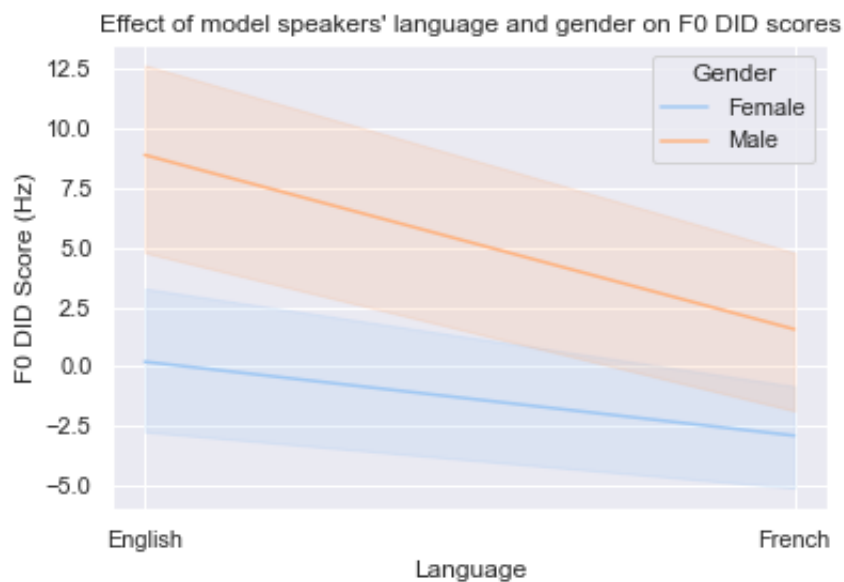


Figure 61: Effect of language and gender on f_0 DID scores. The shaded areas represent the standard deviation of the mean.

Post-hoc comparisons showed significantly more convergence in f_0 towards American English

male model speakers than towards French female ($\beta = -11.83$, $t = -5.15$, $p < .001$), American English female ($\beta = -8.66$, $t = -3.87$, $p < .001$), and French male model speakers ($\beta = 7.35$, $p = 3.23$, $p < .01$). As can be observed in Figure 62, AM_03 ($M = 11.58 \pm 47.11$) received the most alignment on f_0 , followed by AM_02 ($M = 10.51 \pm 45.88$), FM_02 ($M = 4.35 \pm 37.94$), AM_01 ($M = 3.43 \pm 44.82$), FM_05 ($M = 1.58 \pm 38.72$), and AF_01 ($M = 1.54 \pm 36.16$). FF_08 ($M = -3.56 \pm 26.38$) received the most divergence on f_0 , followed by FF_01 ($M = -2.64 \pm 25.93$), FF_02 ($M = -2.35 \pm 21.03$), FM_06 ($M = -0.65 \pm 43.25$), AF_04 ($M = -0.48 \pm 42.34$), and AF_06 ($M = -0.45 \pm 37.62$).

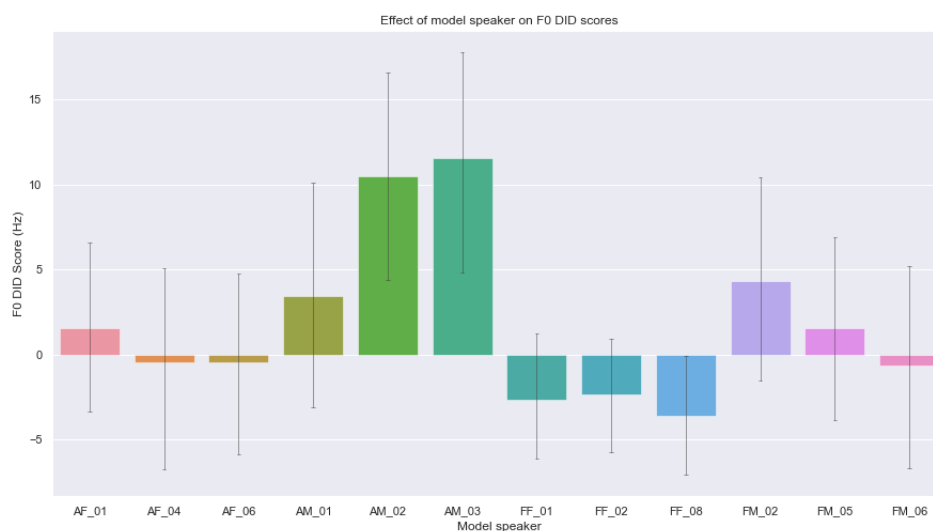


Figure 62: Effect of model speaker on f_0 DID scores. Error bars represent the standard error of the mean.

These results suggest that there was no significant overall alignment on f_0 . However, Language, the interaction between Language and Gender, and Model speaker, all had significant effects on f_0 DID scores. Overall, subjects converged more towards American English than towards French model speakers. Significant gender-based differences were only observed in English, with subjects converging more towards male than towards female model speakers. Male American English model speakers received the most alignment on f_0 . Regarding Model speaker, AM_03 received the most alignment on f_0 and FF_08 received the most divergence.

2.4.3 H1* DID score

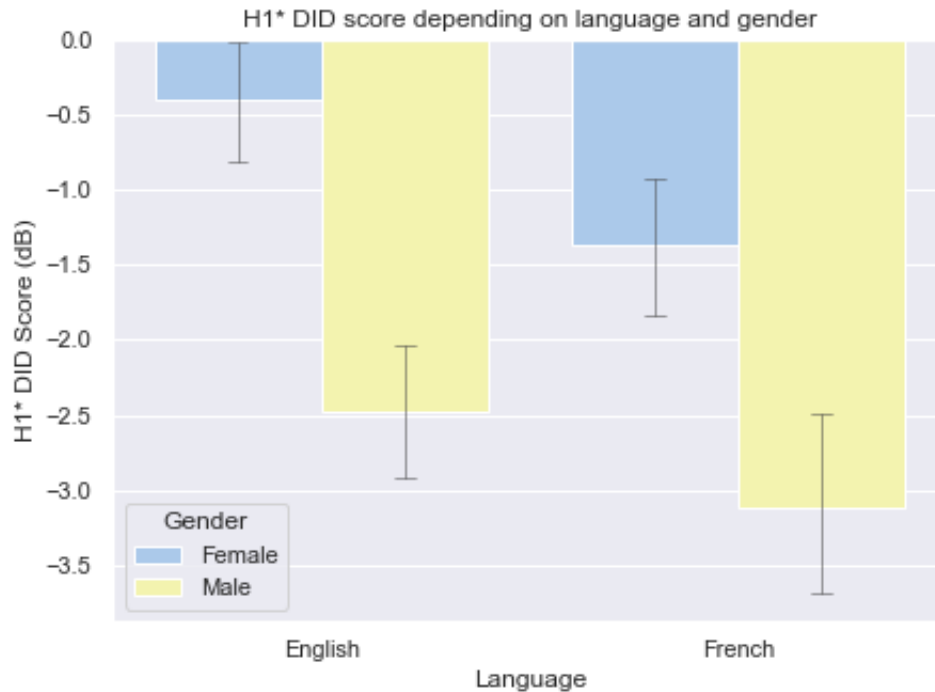


Figure 63: Bar plot representing the mean H1* DID score by language and gender. Error bars represent the standard error of the mean.

Mean H1* DID scores across language and gender are displayed in Figure 63. Subjects have generally diverged from model speakers, with a stronger effect for French model talkers and male model talkers. More divergence effect is observed towards French male model speakers than towards American English male model speakers. The same pattern is observed for female speakers.

Table 31 shows the mean H1* difference values between the reading task and the repetition tasks according to language and model speakers. If we compare these values to the H1* DID values displayed in Figure 63, we can conclude that subjects produced lower H1* values when repeating French male and female model speakers, as well as male American English speakers, leading to divergence effects. They also produced increased H1* values when repeating female

Table 31: Mean H1* difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively).

| Mean H1* difference (dB) | | |
|--------------------------|----------------|-----------------|
| Language | <i>French</i> | \searrow 3.79 |
| | <i>English</i> | \searrow 2.05 |
| Language:Gender | <i>FM</i> | \searrow 5.93 |
| | <i>FF</i> | \searrow 1.65 |
| | <i>AM</i> | \searrow 2.53 |
| | <i>AF</i> | \nearrow 1.59 |

American English speakers, again leading to divergence effects. For convergence to have been observed, subjects should have produced higher H1* values when repeating male and female French model speakers and male American English speakers, but lower H1* values when repeating female American English speakers.

To determine whether Language and the interaction between Language and Gender had a significant effect on H1*, we fitted our control (as described previously) with H1* DID as the dependent measure. The control MEM's intercept was significant ($\beta = -1.81$, $t = -3.94$, $p < .01$), indicating significant overall divergence to H1*. Language ($\beta = -0.85$, $t = -3.8$, $p < .001$; $\chi^2(1) = 14.35$, $p < .001$), the interaction between Language and Gender ($\chi^2(5) = 89.9$, $p < .001$), and Model speaker ($\beta = -1.79$, $t = -2.6$, $p < .05$; $\chi^2(11) = 177.46$, $p < .001$) improved model fit.

As can be observed in Figure 64 and 65, respectively, subjects diverged more from French than from American English model speakers, and more from male than from female model speakers within-language.

Post-hoc comparisons showed significantly less divergence on H1* from female American English than from male American English model speakers ($\beta = 2.07$, $t = 6.74$, $p < .001$), male French model speakers ($\beta = 2.72$, $t = 8.84$, $p < .001$), and female French model speakers ($\beta = 0.95$, $t = 3.06$, $p < .05$), and significantly less from female French model speakers than from male French model speakers ($\beta = 1.77$, $t = 5.6$, $p < .001$). FM_05 received the most divergence (M = -6.17 ± 6.68) on H1*, followed by AM_03 (M = -3.17 ± 5.75), FF_08 (M = -2.4 ± 5.34),

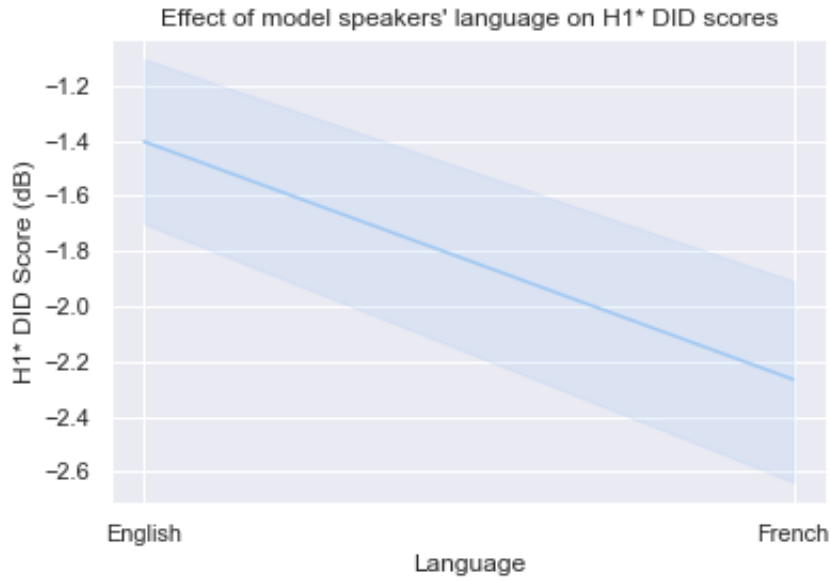


Figure 64: Effect of language on H1* DID scores. The shaded area represents the standard deviation.

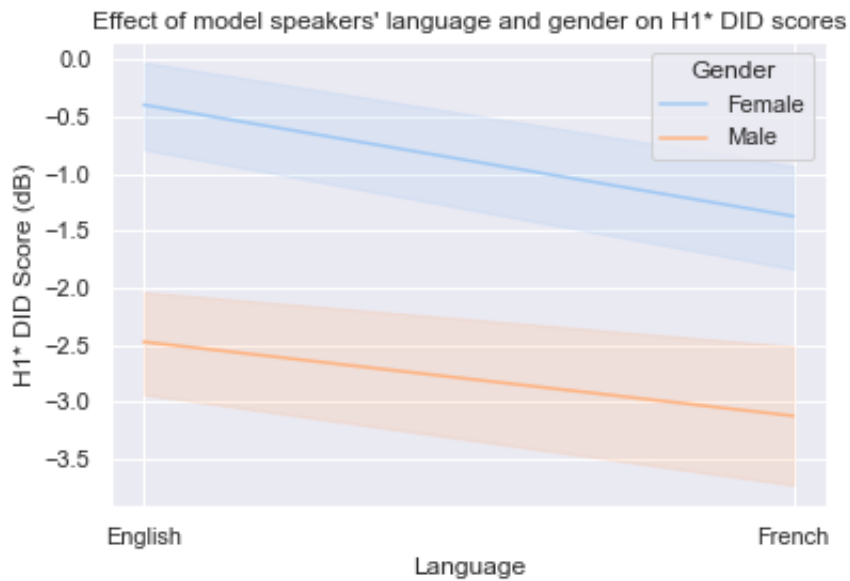


Figure 65: Effect of language and gender on H1* DID scores. The shaded areas represent the standard deviation.

AM_02 ($M = -2.09 \pm 4.81$), AM_01 ($M = -2.04 \pm 5.49$), FM_02 ($M = -1.75 \pm 6.67$), FM_06 ($M = -1.23 \pm 6.99$), FF_01 ($M = -1.09 \pm 4.47$), AF_04 ($M = -1.03 \pm 5.19$), AF_01 ($M = -0.1 \pm 4.99$), and AF_06 ($M = -0.05 \pm 4.28$).



Figure 66: Effect of model speaker on H1* DID scores. Error bars represent the standard error of the mean.

These results suggest that there was overall divergence on H1*. Language, the interaction between Language and Gender, and Model Speaker all had significant effects on H1* DID scores. Overall, subjects diverged more from French than from American English model speakers. Significant gender-based differences were observed within-language, with subjects diverging more from male than from female model speakers, in both French and English. Male French speakers received the most divergence on H1*. Subjects diverged more from FM_05.

2.4.4 H2* DID score

Mean H2* DID scores across language and gender are displayed in Figure 67. As for H1*, subjects have generally diverged from model speakers. Language effect is not very clear, but there is a stronger effect for male model speakers as compared to female ones. More divergence is observed from French male model speakers than from their American English counterparts. Conversely to H1*, there is more divergence effect observed from American English female model speakers than from French ones. Error bars show that some subjects have converged towards female French speakers.

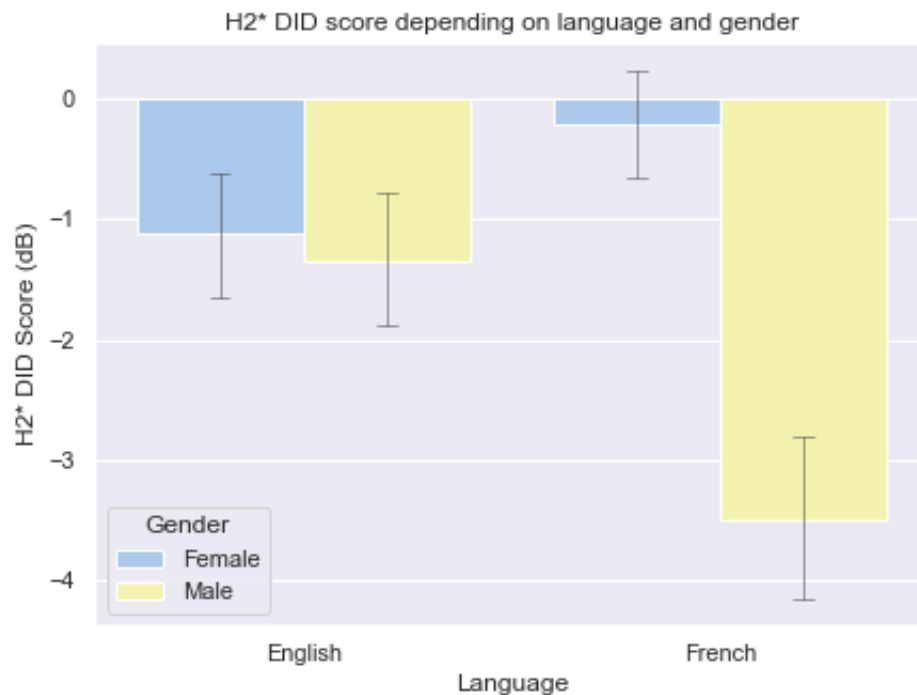


Figure 67: Bar plot representing the mean H2* DID score by language and gender. Error bars represent the standard error of the mean.

Table 32: Mean H2* difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively).

| Mean H2* difference (dB) | | |
|--------------------------|----------------|-----------------|
| Language | <i>French</i> | \searrow 2.46 |
| | <i>English</i> | \searrow 0.84 |
| Language:Gender | <i>FM</i> | \searrow 4.93 |
| | <i>FF</i> | \nearrow 0.02 |
| | <i>AM</i> | \searrow 1.5 |
| | <i>AF</i> | \searrow 0.21 |

Table 32 shows the mean H2* difference values between the reading task and the repetition tasks according to language and model speakers. If we compare these values to the H2* DID values displayed in Figure 67, we can conclude that subjects produced overall lower H2* values when repeating both French and American English model speakers, leading to divergence effects. Lower H2* values were produced when repeating all groups except French female model speakers. For convergence to have been observed, subjects should have produced higher H2* values

when repeating male French model speakers, as well as both female and male American English speakers, but lower H2* values when repeating female French model speakers.

To determine if Language and the interaction between Language and Gender had a significant effect on H2* DID scores, we first fitted our control model (as described previously) with H2* DID as the dependent measure. The control MEM's intercept was significant ($\beta = -1.54$, $t = -2.66$, $p < .05$), indicating significant overall divergence to H2* in that case. Language ($\beta = -0.65$, $t = -2.4$, $p < .05$; $\chi^2(1) = 5.74$, $p < .05$), the interaction between Language and Gender ($\beta = -3.08$, $t = -5.77$, $p < .001$; $\chi^2(5) = 79.27$, $p < .001$), and Model speaker ($\beta = -1.52$, $t = -2.1$, $p < .05$; $\chi^2(11) = 98.29$, $p < .001$) improved model fit.

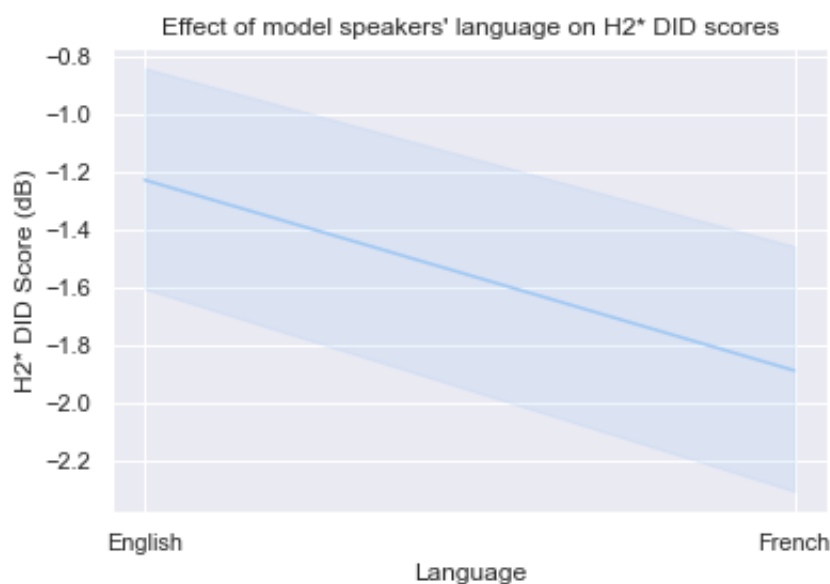


Figure 68: Effect of language on H2* DID scores. The shaded area represents the standard deviation.

As can be observed in Figure 68, subjects diverged significantly more from French than from American English model speakers. *Post-hoc* comparisons showed significantly more divergence from French male than from French female ($\beta = 3.31$, $t = 8.62$, $p < .001$), American English female ($\beta = 2.38$, $t = 6.37$, $p < .001$), and American English male model speakers ($\beta = 2.16$, $t = 5.69$, $p < .001$), as illustrated in Figure 69.

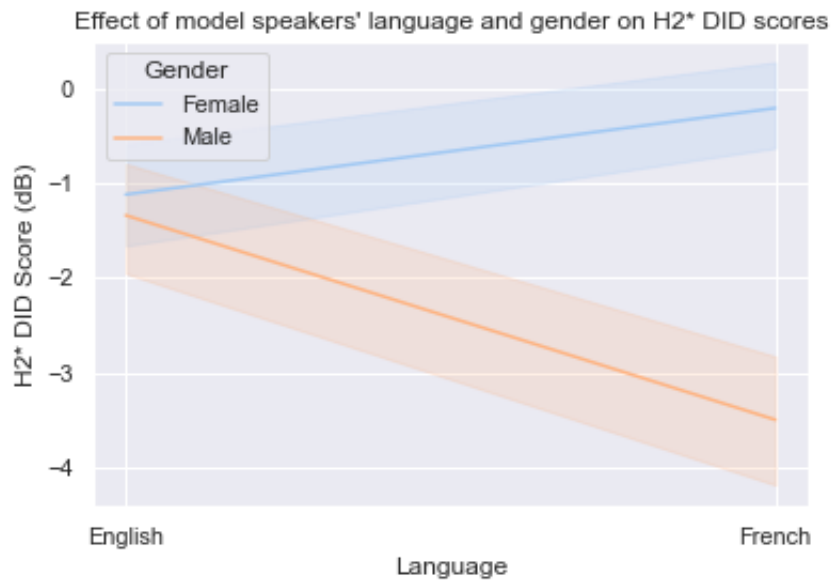


Figure 69: Effect of language and gender on H2* DID scores. The shaded area represents the standard deviation.

As can be seen in Figure 70, subjects only converged on H2* towards FF_02 ($M = 0.8 \pm 5.45$). FM_05 received the most divergence on H2* ($M = -5.42 \pm 8.41$), followed by FM_02 ($M = -2.81 \pm 7.45$), AF_04 ($M = -2.64 \pm 7.12$), FM_06 ($M = -2.17 \pm 7.84$), AM_02 ($M = -1.94 \pm 5.73$), AM_01 ($M = -1.22 \pm 7.3$), FF_01 ($M = -1.2 \pm 4.78$), AM_03 ($M = -0.8 \pm 7.58$), AF_06 ($M = -0.54 \pm 6.29$), AF_01 ($M = -0.17 \pm 5.82$), FF_08 ($M = -0.13 \pm 5.44$),

These results suggest that there was significant overall divergence on H2*. Language, the interaction between Language and Gender, and Model Speaker had an effect on H2* DID scores. Overall, subjects diverged more from French model speakers than from American English ones. Significant gender-based differences were only observed in French, with subjects diverging more from male than from female model speakers. Female French speakers received the least divergence on H2*. Regarding Model speaker, subjects only converged towards FF_02 and FM_05 received the most divergence.

2.4.5 H1*-H2* DID score

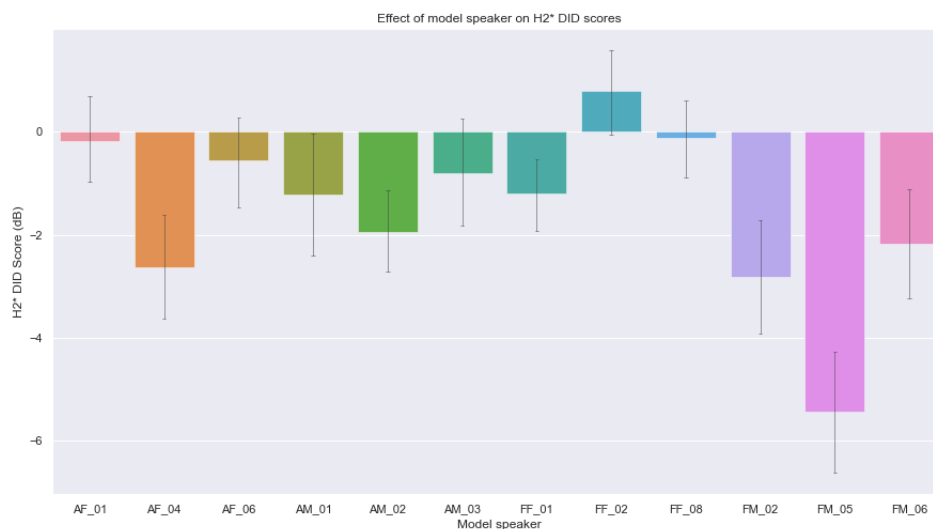


Figure 70: Effect of model speaker on H2* DID scores. Error bars represent the standard deviation of the mean.

Mean H1*-H2* DID scores across language and gender are displayed in Figure 71. Overall, there is a strong trend towards divergence from French model speakers and a gender-dependent weak trend towards either convergence or divergence towards/from American English model speakers. Subjects have diverged slightly more from female French model speakers than from male French ones on H1*-H2*. A weaker divergence effect is also observed towards female American English model speakers while learners have generally converged towards male American English model speakers. Error bars indicate that some subjects have also converged towards female American English speakers and some have diverged from male American English speakers.

Table 33 shows the mean H1*-H2* difference values between the reading task and the repetition tasks according to language and model speakers. If we compare these values to the H1*-H2* DID values displayed in Figure 71, we can conclude that subjects produced overall lower H1*-H2* values when repeating all groups, hence creakier vowels. For convergence to have been observed, subjects should have overall produced increased H1*-H2* values when repeating male and female French model speakers, and decreased H1*-H2* values when repeating male American English speakers. It was not very clear whether they should have produced more increased or decreased

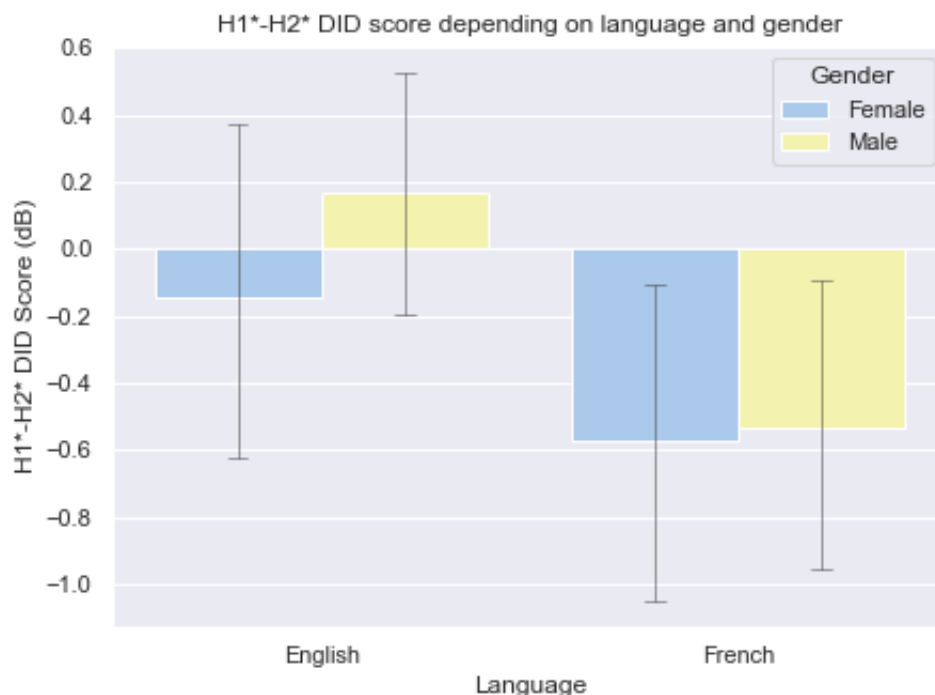


Figure 71: Bar plot representing the mean H1*-H2* DID score by language and gender. Error bars represent the standard error of the mean.

Table 33: Mean H1*-H2* difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively).

| Mean H1*-H2* difference (dB) | | |
|------------------------------|----------------|-----------------|
| Language | <i>French</i> | \searrow 1.34 |
| | <i>English</i> | \searrow 1.21 |
| Language:Gender | <i>FM</i> | \searrow 1.0 |
| | <i>FF</i> | \searrow 1.67 |
| | <i>AM</i> | \searrow 1.03 |
| | <i>AF</i> | \searrow 1.39 |

H1*-H2* values to converge towards female American English speakers, so we compared mean H1*-H2* values from the reading and repetition tasks to that obtained for female American English model speakers. We concluded that more subjects should have produced lower H1*-H2* values for overall convergence to be observed towards female American English models.

To determine if Language and the interaction between Language Gender had a significant

effect on H1*-H2* DID scores, we first fitted our control model (as described previously) with H1*-H2* DID as the dependent measure. The control MEM's intercept was non-significant ($\beta = -0.26$, $t = -1.78$, $p = .09$), indicating no significant overall accommodation⁶¹ to H1*-H2*. Language ($\beta = -0.56$, $t = -2.54$, $p < .05$; $\chi^2(1) = 6.46$, $p < .05$) and Model speaker ($\chi^2(11) = 10.66$, $p < .01$) improved model fit. The interaction between Language and Gender ($\beta = -0.28$, $t = -0.64$, $p = .52$; $\chi^2(5) = 7.54$, $p = .06$) did not. Subjects diverged more from French model speakers than from American English model speakers, as can be observed in Figure 72.

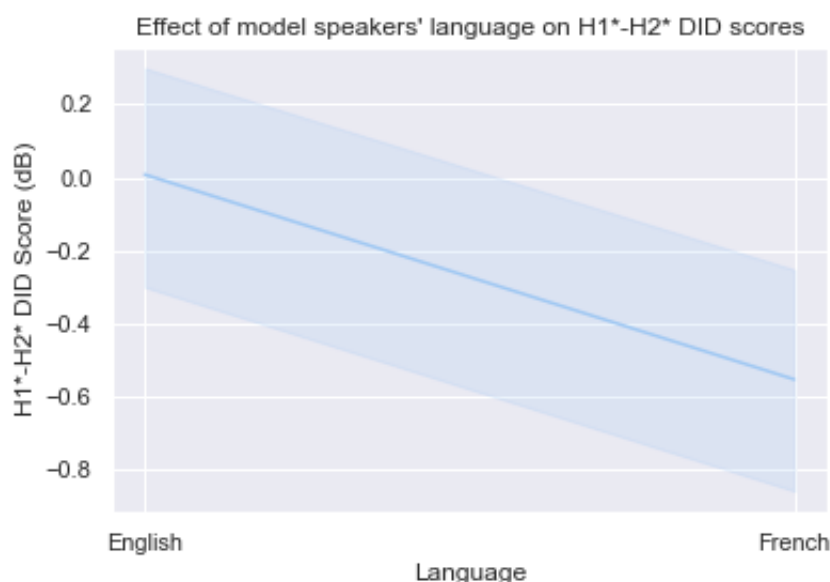


Figure 72: Effect of language on H1*-H2* DID scores. The shaded area represents the standard deviation.

As can be seen in Figure 73, AF_06 ($M = 0.8 \pm 5.99$) received the most convergence on H1*-H2*, followed by AM_01 ($M = 0.73 \pm 4.12$), and FF_02 ($M = 0.51 \pm 4.6$). Subjects diverged from all other model speakers, with FF_01 ($M = -1.54 \pm 5.38$) receiving the most divergence followed by FM_02 ($M = -0.97 \pm 4.79$), AF_01 ($M = -0.92 \pm 6.05$), FF_08 ($M = -0.57 \pm 4.99$), FM_05 ($M = -0.38 \pm 5.06$), FM_06 ($M = -0.35 \pm 5.43$), AF_04 ($M = -0.31 \pm 6.18$), AM_02 ($M = -0.08 \pm 4.23$), and AM_03 ($M = -0.02 \pm 4.69$).

⁶¹We use 'accommodation' here because it is not clear whether the accommodation pattern is divergence or convergence

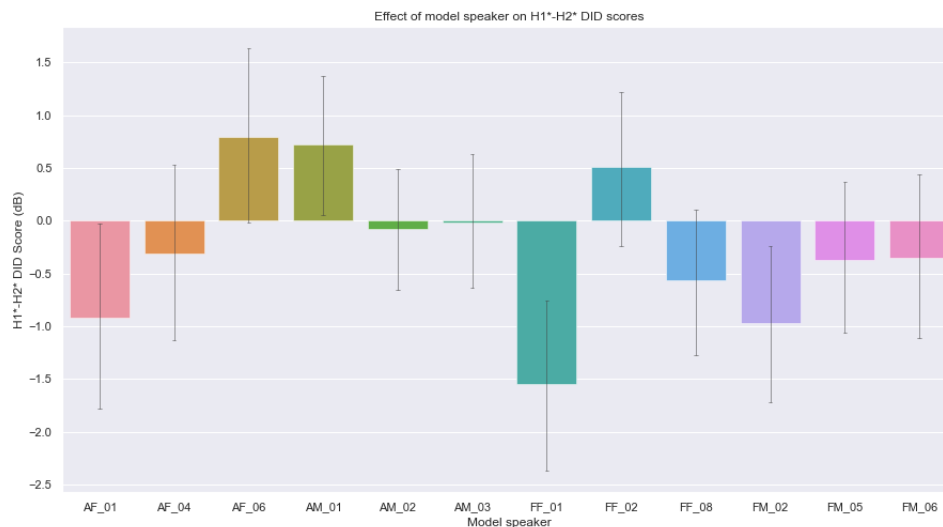


Figure 73: Effect of model speaker on H1*-H2* DID scores. Error bars represent the standard deviation of the mean.

These results suggest that there was no significant overall alignment on H1*-H2*. The interaction between Language and Gender did not result in any effect. Effects were significant for Language and for Model speaker. Overall, subjects diverged more from French model speakers than from American English ones on H2*. Regarding Model speaker, AF_06 received the most convergence on H1*-H2* and FF_01 the most divergence.

2.4.6 H1*-A1* DID score

Mean H1*-A1* DID scores across language and gender are displayed in Figure 74. Subjects have generally diverged from model speakers, with a stronger effect for French and male model speakers. Divergence effects are stronger for French male model speakers than for American English male model speakers. The same pattern is observed for female speakers. Error bars indicate that some subjects have converged towards both French and American English female model speakers.

Table 34 shows the mean H1*-A1* difference values between the reading task and the repetition tasks according to language and model speakers. If we compare these values to the H1*-A1*

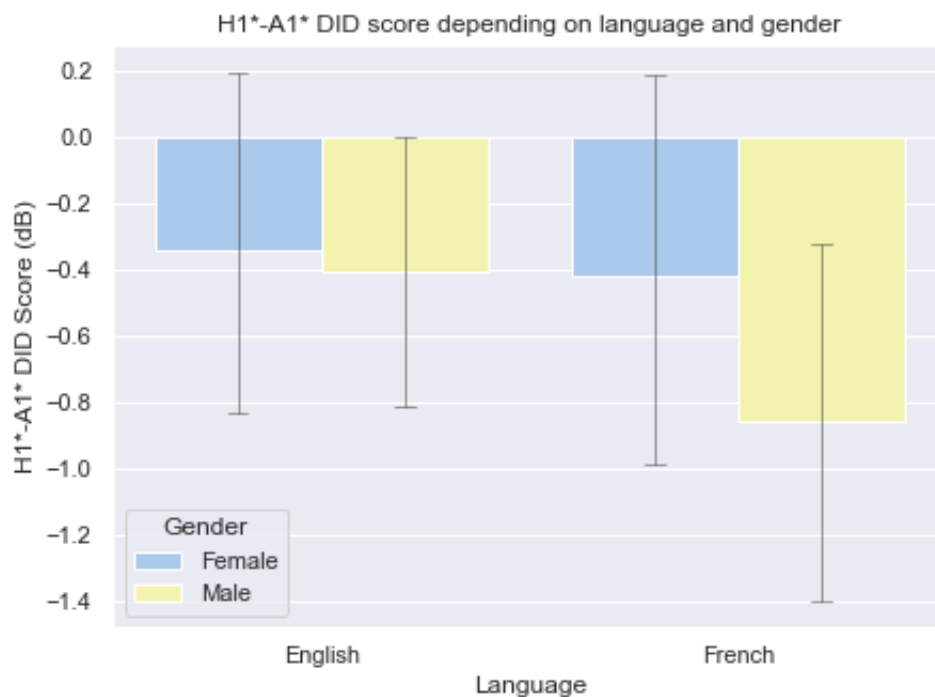


Figure 74: Bar plot representing the mean H1*-A1* DID score by language and gender. Error bars represent the standard error of the mean.

Table 34: Mean H1*-A1* difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively).

| Mean H1*-A1* difference (dB) | | |
|------------------------------|----------------|-----------------|
| Language | <i>French</i> | \searrow 1.17 |
| | <i>English</i> | \nearrow 0.6 |
| Language:Gender | <i>FM</i> | \searrow 0.79 |
| | <i>FF</i> | \searrow 1.55 |
| | <i>AM</i> | \searrow 0.04 |
| | <i>AF</i> | \nearrow 1.19 |

DID values displayed in Figure 74, we can conclude that subjects produced overall lower H1*-A1* values when repeating French model speakers and higher H1*-A1* values when repeating American English model speakers. If we now take gender into consideration, we can see that subjects produced decreased H1*-A1* values when repeating both male and female French speakers, as well as when repeating American English male model speakers. They produced increased H1*-

A1* values when repeating American English female speakers. For convergence effects to have been observed, more subjects should have produced increased H1*-A1* values when repeating all groups.

To determine if Language and the interaction between Language and Gender had a significant effect on H1*-A1* DID scores, we first fitted our control model (as described previously) with H1*-A1* DID as the dependent measure. The control MEM's intercept was significant ($\beta = -0.5$, $t = -2.51$, $p < .05$), indicating significant overall divergence to H1*-A1* in that case. Language ($\beta = -0.27$, $t = -1.05$, $p = .03$; $\chi^2(1) = 1.09$, $p = .03$), the interaction between Language and Gender ($\beta = -0.38$, $t = -0.72$, $p = .47$; $\chi^2(5) = 2.49$, $p = .48$), and Model speaker ($\chi^2(11) = 1.96$, $p = .16$) did not improve model fit.

These results suggest that there was significant overall divergence on H1-A1*. Language, the interaction of Language and Gender, or Model Speaker did not result in any significant effect.

2.4.7 H1*-A2* DID score

Mean H1*-A2* DID scores across language and gender are displayed in Figure 75. There is an overall trend towards divergence across language, with a stronger effect for male than for female model speakers. Divergence effects are stronger for American English male model speakers than for French ones. There is a very weak trend towards convergence for female American English model speaker only. Error bars indicate that some subjects have also converged towards French female model speakers and others have diverged from American English female model speakers.

Table 35 shows the mean H1*-A2* difference values between the reading task and the repetition tasks according to language and model speakers. If we compare these values to the H1*-A2* DID values displayed in Figure 75, we can conclude that subjects produced overall lower H1*-A2* values when repeating both French and American English model speakers, and actually when repeating all groups when splitting between-gender. For convergence effects to have been observed, more subjects should have produced increased H1*-A2* values when repeating all groups.

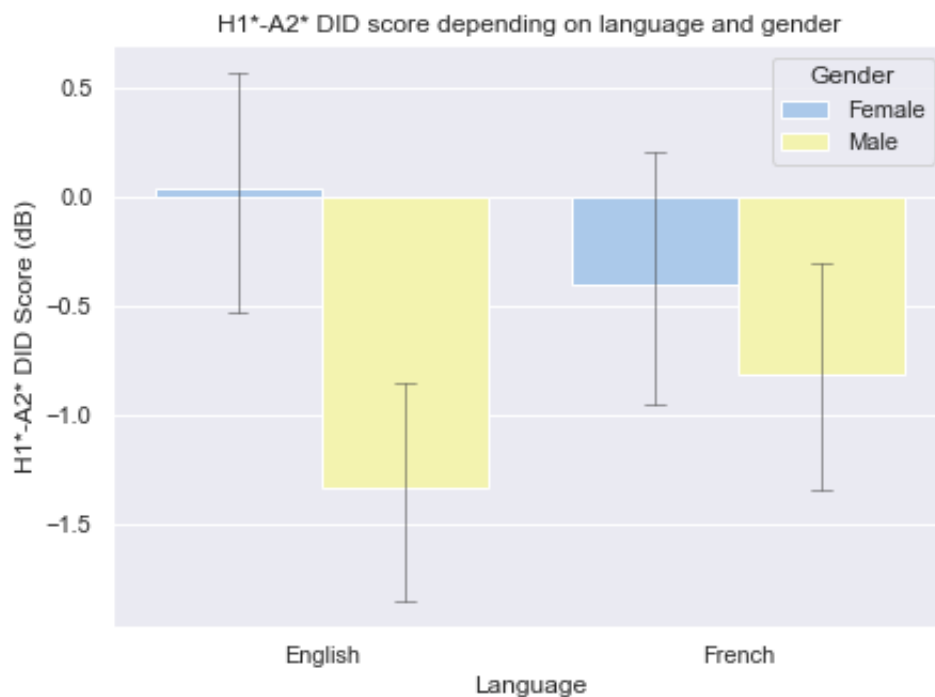


Figure 75: Bar plot representing the mean H1*-A2* DID score by language and gender. Error bars represent the standard error of the mean.

Table 35: Mean H1*-A2* difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively).

| Mean H1*-A2* difference (dB) | | |
|------------------------------|----------------|-----------------|
| Language | <i>French</i> | \searrow 1.37 |
| | <i>English</i> | \searrow 1.11 |
| Language:Gender | <i>FM</i> | \searrow 1.34 |
| | <i>FF</i> | \searrow 1.41 |
| | <i>AM</i> | \searrow 0.88 |
| | <i>AF</i> | \searrow 1.34 |

To determine if Language and the interaction between Language and Gender had a significant effect on H1*-A2* DID scores, we first fitted our control model (as described previously) with H1*-A2* DID as the dependent measure. The control MEM's intercept was significant ($\beta = -0.62$, $t = -3.2$, $p < .01$), indicating significant overall divergence to H1*-A2*. Language did not improve model fit ($\beta = 0.01$, $t = 0.05$, $p = .96$; $\chi^2(1) = 0.003$, $p = .96$). However, the interaction

between Language and Gender ($\chi^2(5) = 14.04, p < .01$), as well as Model speaker ($\beta = -0.61, t = -2.21, p < .05; \chi^2(11) = 11.85, p < .001$) improved the model fit.

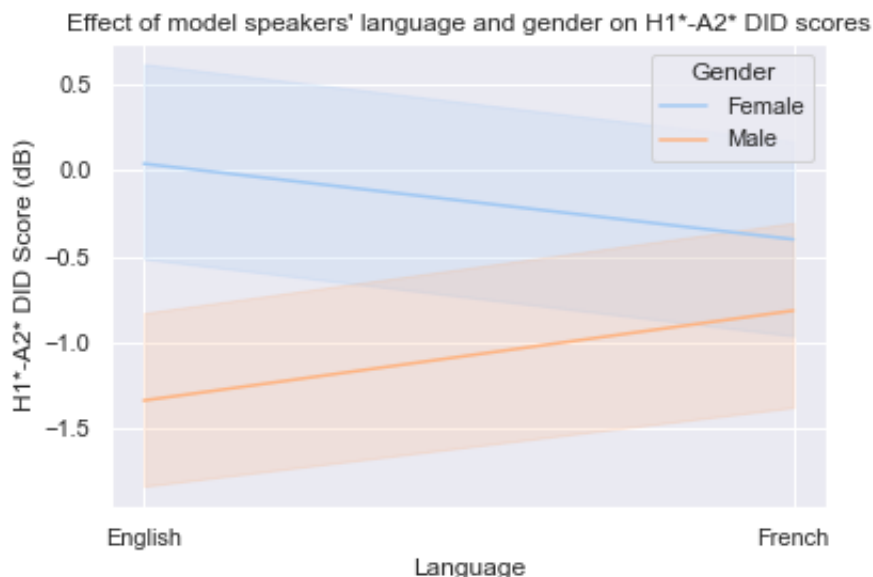


Figure 76: Effect of language and gender on H1*-A2* DID scores. The shaded areas represent the standard deviation.

Post-hoc comparisons showed significantly more divergence from male American English than from female American English model speakers ($\beta = 1.38, t = 3.6, p < .01$), as illustrated in Figure 76.

As can be seen in Figure 77, subjects only converged towards AF_01 ($M = 1.29 \pm 6.88$) on H1*-A2*. They diverged from all other model speakers, with AM_03 receiving the most divergence ($M = -2.16 \pm 7.83$), followed by FM_05 ($M = -1.47 \pm 6.00$), AM_02 ($M = -1.12 \pm 4.59$), AF_06 ($M = -0.99 \pm 6.88$), FF_01 ($M = -0.69 \pm 7.23$), AM_01 ($M = -0.64 \pm 4.05$), FM_06 ($M = -0.63 \pm 6.71$), FF_08 ($M = -0.37 \pm 7.24$), AF_04 ($M = -0.17 \pm 6.67$), FM_02 ($M = -0.16 \pm 5.60$), and FF_02 ($M = -0.11 \pm 6.14$).

These results suggest that there was significant overall divergence on H1*-A2*. H1*-A2* divergence did not vary by language. Effects were significant for the interaction between Language

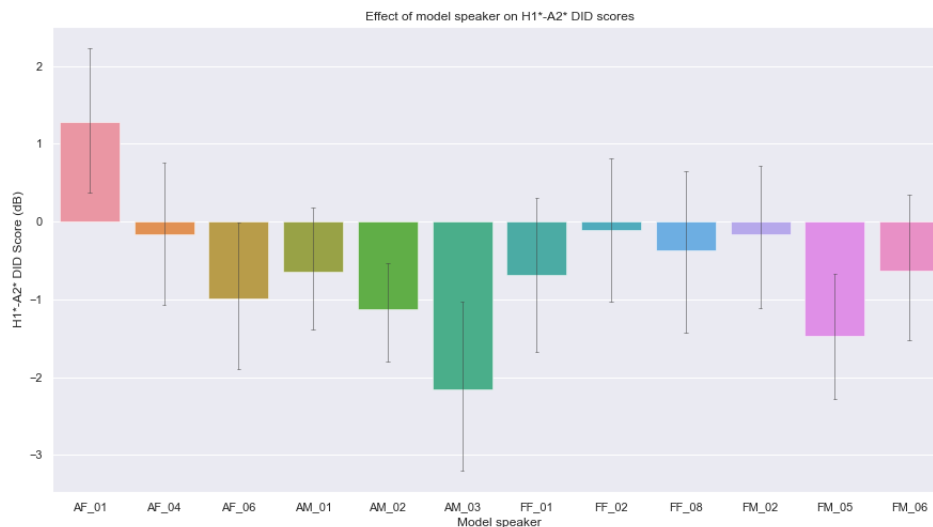


Figure 77: Effect of model speaker on H1*-A2* DID scores. Error bars represent the standard deviation of the mean.

and Gender, as well as for Model speaker. Overall, subjects diverged more from male than from female speakers. Gender-based differences were only significant in American English, with more divergence observed from male than from female model speakers, for which overall maintenance, or very little convergence was observed. Female American English model speakers received the least divergence on H1*-A2*. Regarding Model speaker, subjects only converged towards AF_01 and AM_03 received the most divergence on H1*-A2*.

2.4.8 H1*-A3* DID score

Mean H1*-A3* DID scores across language and gender are displayed in Figure 78. As for H1*-A2*, there is an overall trend towards divergence across language, with a stronger effect for male than for female speakers, as well as for French model speakers. Divergence effects are stronger for French male model speakers than for American English male ones in this case. Convergence is observed towards female American English model speakers only.

Table 36 shows the mean H1*-A3* difference values between the reading task and the repetition tasks according to language and model speakers. Subjects produced overall lower H1*-A3*

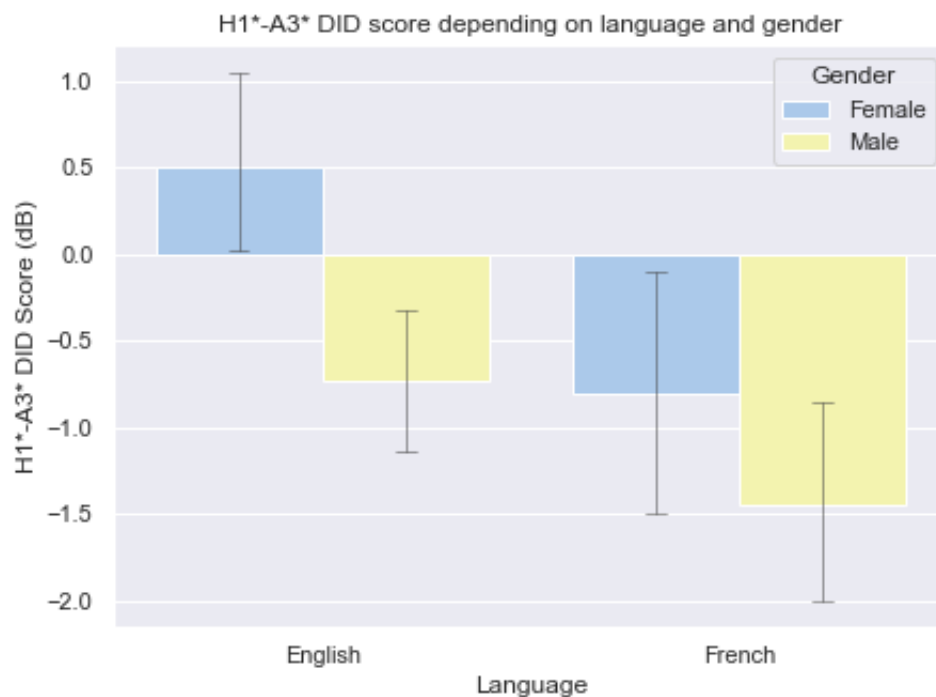


Figure 78: Bar plot representing the mean H1*-A3* DID score by language and gender. Error bars represent the standard error of the mean.

Table 36: Mean H1*-A3* difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively).

| Mean H1*-A3* difference (dB) | | |
|------------------------------|----------------|-------------------|
| Language | <i>French</i> | \searrow 1.71 |
| | <i>English</i> | \nearrow 0.0001 |
| Language:Gender | <i>FM</i> | \searrow 1.93 |
| | <i>FF</i> | \searrow 1.49 |
| | <i>AM</i> | \searrow 0.59 |
| | <i>AF</i> | \nearrow 0.56 |

values when repeating both French model speakers, and very slightly higher H1*-A3* values when repeating American English model speakers. If we break down the data across gender, we can observe decreased values when repeating both male and female French model speakers, as well as when repeating male American English ones. Subjects only produced increased H1*-A3* values when repeating female American English model speakers. If we now compare these values

to the H1*-A3* DID values displayed in Figure 78, we can conclude that, for convergence effects to have been observed towards male and female French model speakers and male American English ones, subjects should have produced lower H1*-A3* values than what they did during the reading task.

To determine if Language and the interaction between Language Gender had a significant effect on H1*-A3* DID scores, we first fitted our control model (as described previously) with H1*-A3* DID as the dependent measure. The control MEM's intercept was significant ($\beta = -0.6$, $t = -2.67$, $p < .05$), indicating significant overall divergence to H1*-A3*. Language ($\beta = -1.04$, $t = -3.6$, $p < .001$; $\chi^2(1) = 12.92$, $p < .001$), the interaction between Language and Gender ($\chi^2(5) = 24.9$, $p < .001$), and Model speaker ($\chi^2(11) = 32.96$, $p < .001$) improved model fit.

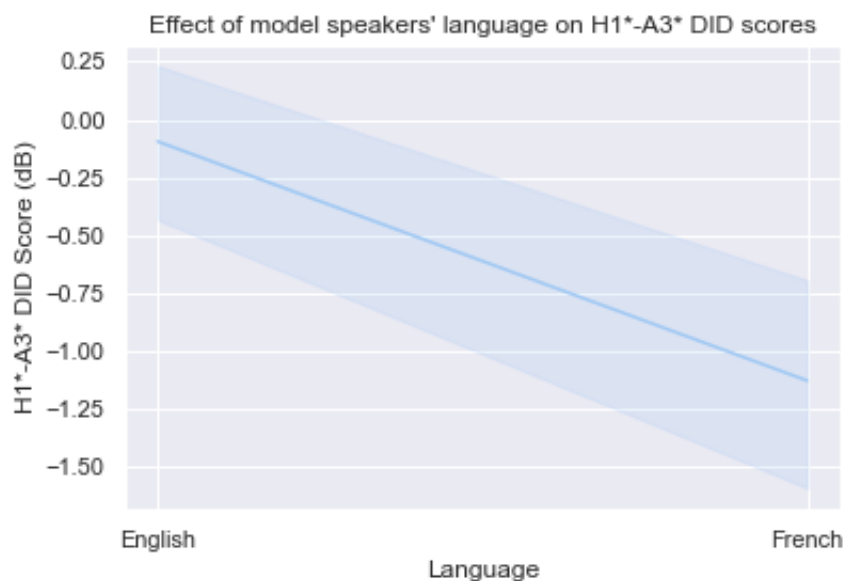


Figure 79: Effect of language on H1*-A3* DID scores. The shaded area represents the standard deviation.

We can observe in Figure 79 that subjects diverged significantly more from French model speakers than from American English ones. *Post-hoc* comparisons showed that differences in H1*-A3* DID scores were significant for female American English model speakers as compared to French male ($\beta = 1.95$, $t = 4.86$, $p < .001$), French female ($\beta = 1.32$, $t = 3.24$, $p < .01$), and American English male model speakers ($\beta = 1.24$, $t = 3.1$, $p < .05$), as illustrated in Figure 80.

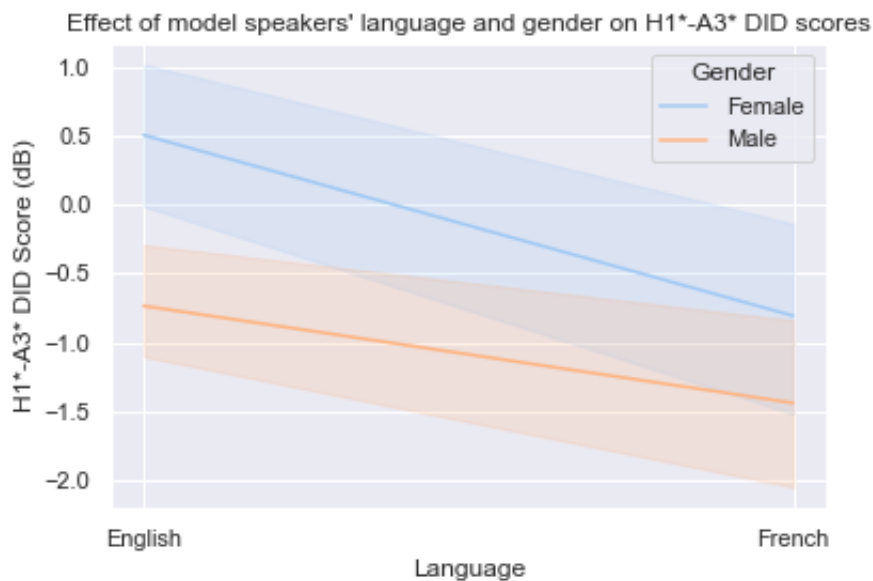


Figure 80: Effect of language and gender on H1*-A3* DID scores. The shaded areas represent the standard deviation.

As can be seen in Figure 83, subjects only converged towards AF_01 ($M = 1.72 \pm 6.66$) and AF_04 ($M = 0.91 \pm 6.21$). Subjects diverged from all other model speakers, with FM_05 receiving the most divergence ($M = -3.04 \pm 8.52$), followed by AM_03 ($M = -1.19 \pm 5.56$), AF_06 ($M = -1.09 \pm 6.24$), AM_01 ($M = -1.01 \pm 5.04$), FF_02 ($M = -0.96 \pm 7.8$), FM_02 ($M = -0.92 \pm 5.96$), FF_08 ($M = -0.79 \pm 7.95$), FF_01 ($M = -0.69 \pm 8.28$), FM_06 ($M = -0.23 \pm 7.14$), and AM_02 ($M = -0.09 \pm 4.5$).

These results suggest that there was significant overall divergence on H1*-A3*. Language, the interaction between Language and Gender, as well as Model speaker, had an effect on H1*-A3* DID scores. Overall, there was more divergence observed from French model speakers than from American English model speakers, and more from male than from female model speakers. Gender-based differences were only significant in English, with subjects diverging from male and converging towards female model speakers. Female American English model speakers received the most alignment on H1*-A3* and FM_05 the most divergence.

2.4.9 H2*-H4* DID score

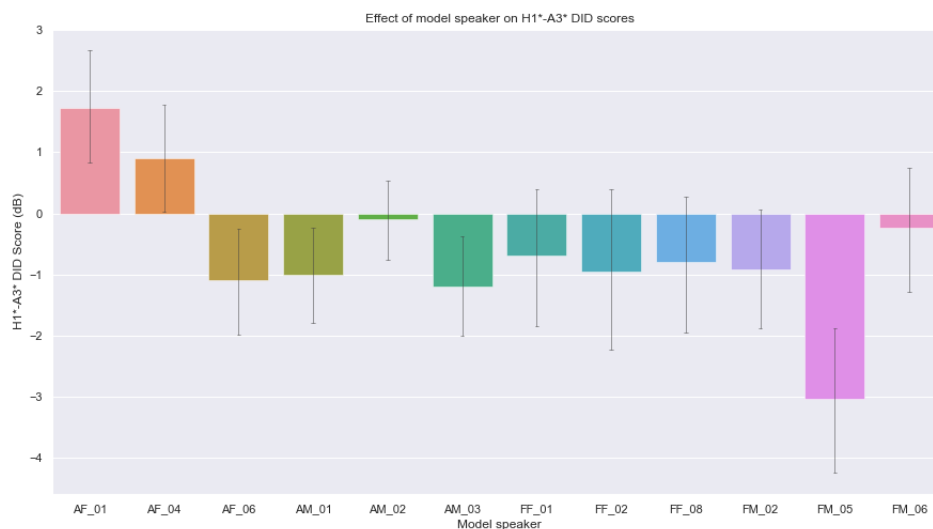


Figure 81: Effect of model speaker on H1*-A3* DID scores. Error bars represent the standard deviation of the mean.

Mean H2*-H4* DID scores across language and gender are displayed in Figure 82. Overall more divergence effects are observed, with a stronger effect for male and American English model speakers. Divergence effects are stronger for French male model speakers than for their American English counterparts. Convergence is only observed towards French female model speakers. Error bars indicate that some subjects have also converged towards both male and female speakers of American English, as well as towards French male model speakers. Some subjects have also diverged from female French speakers.

Table 37: Mean H2*-H4* difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively).

| Mean H2*-H4* difference (dB) | | |
|------------------------------|----------------|-----------------|
| Language | <i>French</i> | \searrow 0.68 |
| | <i>English</i> | \searrow 0.08 |
| Language:Gender | <i>FM</i> | \searrow 0.83 |
| | <i>FF</i> | \searrow 0.53 |
| | <i>AM</i> | \searrow 0.17 |
| | <i>AF</i> | \nearrow 0.32 |

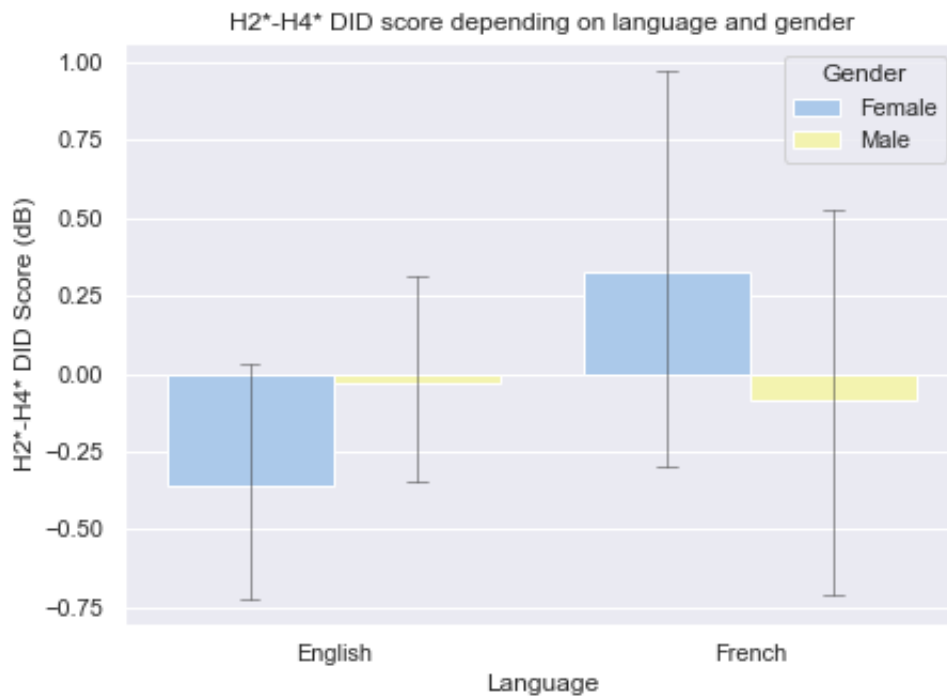


Figure 82: Bar plot representing the mean H2*-H4* DID score by language and gender. Error bars represent the standard error of the mean.

Table 37 shows the mean H2*-H4* difference values between the reading task and the repetition tasks according to language and model speakers. Subjects produced overall lower H2*-H4* values when repeating both French and American English model speakers. If we break down the data across gender, we can observe decreased values when repeating both male and female French model speakers, as well as when repeating male American English ones. Subjects only produced increased H2*-H4* values when repeating female American English model speakers. If we now compare these values to the H2*-H4* DID values displayed in Figure 82, we can conclude that, for overall convergence effects to have been observed towards both French and American English male model speakers, more subjects should have produced increased H2*-H4* values when repeating those groups as compared to that in the reading task. For convergence to be observed towards female American English model speakers, subjects should have produced decreased H2*-H4* values when repeating them.

To determine if Language and the interaction between Language Gender had a significant

effect on H2*-H4* DID scores, we first fitted out control model (as described previously) with H2*-H4* DID as the dependent measure. The control MEM's intercept was non-significant ($\beta = -0.05$, $t = -0.28$, $p = .78$), indicating no significant overall accommodation to H2*-H4*. Language ($\beta = 0.31$, $t = 1.24$, $p = .22$; $\chi^2(1) = 1.53$, $p = .22$), and the interaction between Language and Gender ($\beta = -0.73$, $t = -1.45$, $p = .15$; $\chi^2(5) = 3.64$, $p = .3$) did not not improve model fit, but Model speaker did ($\chi^2(11) = 3.98$, $p < .05$).

As can be seen in Figure 83, subjects converged more towards AF_01 ($M = 0.62 \pm 4.5$) on H2*-H4*, then towards FF_02 ($M = 0.49 \pm 7.41$) FM_06 ($M = 0.48 \pm 7.58$), AM_03 ($M = 0.45 \pm 4.13$), FF_08 ($M = 0.34 \pm 8.31$), FM_02 ($M = 0.31 \pm 6.29$), FF_01 ($M = 0.17 \pm 7.48$), and AM_01 ($M = 0.04 \pm 4.07$). AF_06 ($M = -1.32 \pm 4.67$) received the most divergence, followed by FM_05 ($M = -0.94 \pm 6.91$), AM_02 ($M = -0.59 \pm 4.06$), and AF_04 ($M = -0.37 \pm 4.42$).

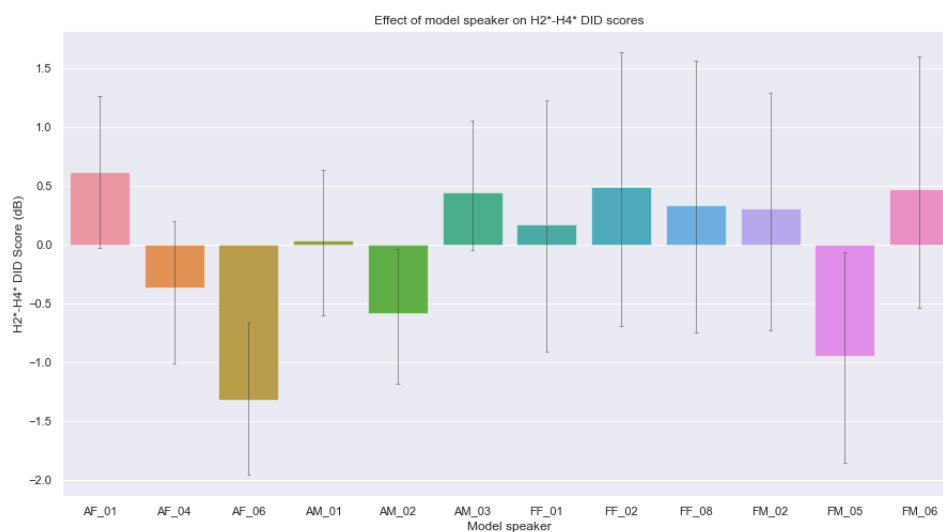


Figure 83: Effect of model speaker on H2*-H4* DID scores. Error bars represent the standard deviation of the mean.

These results suggest that there was no overall convergence effects observed on H2*-H4*. Language and the interaction between Language and Gender did not result in significant effect on H2*-H4* DID scores. There were only more effects observed towards particular model speakers, with AF_01 receiving the most convergence and AF_06 the most divergence.

2.4.10 CPP DID score

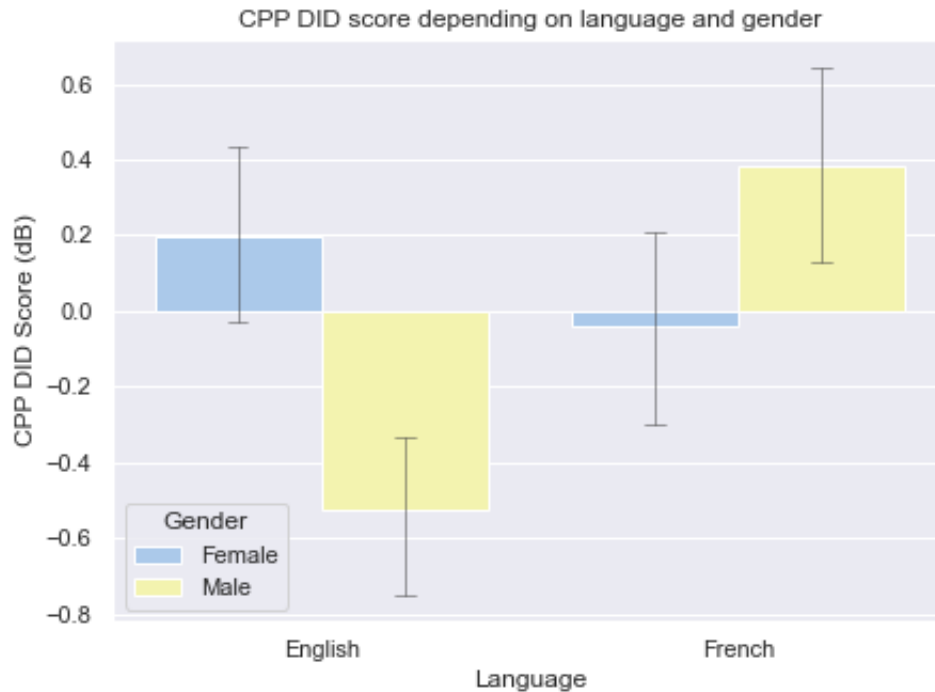


Figure 84: Bar plot representing the mean CPP DID score by language and gender. Error bars represent the standard error of the mean.

Mean CPP DID scores across language and gender are displayed in Figure 84. The pattern is here completely reversed across language. Subjects have overall converged towards female speakers of American English but diverged from French ones, and converged towards male speakers of French but diverged from American English ones. Error bars indicate that some subjects have also converged towards French female model speakers and diverged from American English female model speakers.

Table 38 shows the mean CPP difference values between the reading task and the repetition tasks according to language and model speakers. Subjects produced overall lower CPP values when repeating both French and American English model speakers, and actually when repeating all groups. If we now compare these values to the CPP DID values displayed in Figure 84, we

Table 38: Mean CPP difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively).

| Mean CPP difference (dB) | | |
|--------------------------|----------------|-----------------|
| Language | <i>French</i> | \searrow 1.4 |
| | <i>English</i> | \searrow 1.05 |
| Language:Gender | <i>FM</i> | \searrow 2.31 |
| | <i>FF</i> | \searrow 0.49 |
| | <i>AM</i> | \searrow 1.17 |
| | <i>AF</i> | \searrow 0.94 |

can conclude that, for convergence effects to have been observed towards male American English model speakers and female French model speakers, subjects should have produced increased CPP values as compared to that in the reading task.

To determine if Language and the interaction between Language Gender had a significant effect on CPP DID scores, we first fitted our control model (as described previously) with CPP DID as the dependent measure. The control MEM's intercept was non-significant ($\beta = 0.001$, $t = 0.06$, $p = .95$), indicating no significant overall accommodation to CPP. Language ($\beta = 0.33$, $t = 2.64$, $p < .01$; $\chi^2(1) = 6.98$, $p < .01$), the interaction between Language and Gender ($\beta = 1.15$, $t = 4.65$, $p < .001$; $\chi^2(5) = 30.43$, $p < .001$) and Model speaker ($\chi^2(1) = 102.31$, $p < .001$) improved model fit.

We can observe in Figure 85 that there was significantly more convergence observed towards French model speakers and more divergence from American English speakers. *Post-hoc* comparisons showed that differences in CPP DID scores were significant for American English male model speakers as compared to French male ($\beta = -0.91$, $t = -5.19$, $p < .001$), French female ($\beta = 0.486$, $t = 2.75$, $p < .05$), and American English female model speakers ($\beta = 0.727$, $t = 4.21$, $p < .001$), as illustrated in Figure 86.

As can be seen in Figure 87, subjects converged more towards AF_04 ($M = 1.35 \pm 2.95$) on CPP, then towards FM_05 ($M = 0.89 \pm 2.49$), FF_08 ($M = 0.52 \pm 2.7$), FM_02 ($M = 0.48 \pm 3.36$), and FF_01 ($M = 0.43 \pm 2.61$). Subjects diverged from all other model speakers,

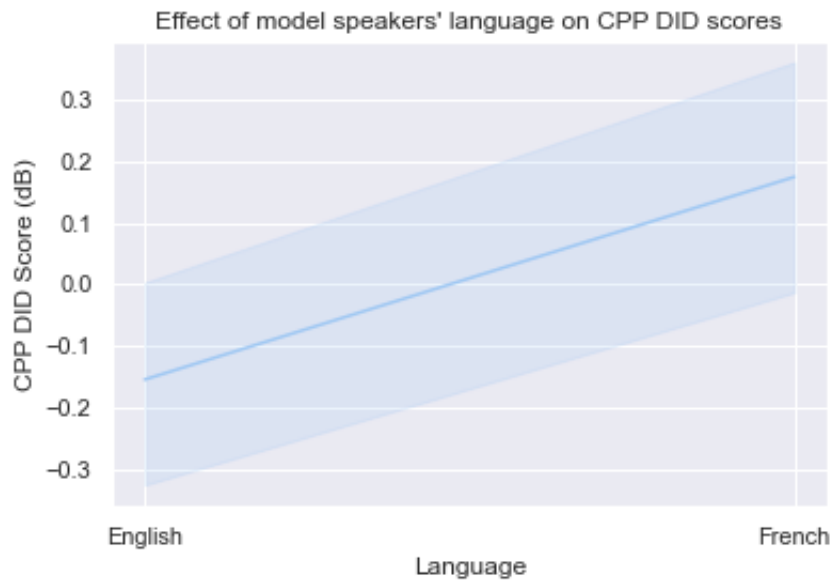


Figure 85: Effect of language on CPP* DID scores. The shaded area represents the standard deviation.

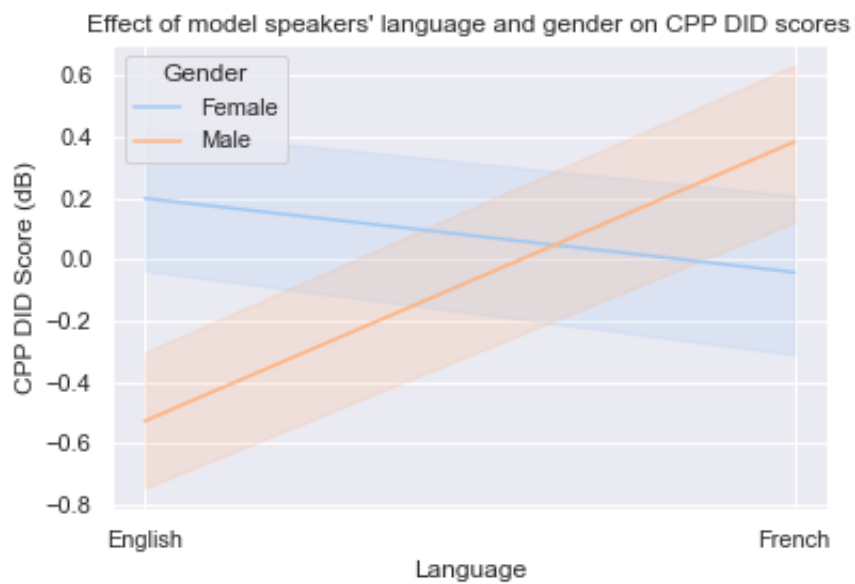


Figure 86: Effect of language and gender on CPP DID scores. The shaded areas represent the standard deviation.

with FF_02 ($M = -1.27 \pm 3.55$) receiving the most divergence, followed by AM_02 ($M = -1.27 \pm 3.55$), AF_06 ($M = -0.42 \pm 3.04$), AF_01 ($M = -0.34 \pm 2.53$), AM_03 ($M = -0.3 \pm 2.78$), FM_06 ($M = -0.21 \pm 3.3$), and AM_01 ($M = -0.12 \pm 2.41$).

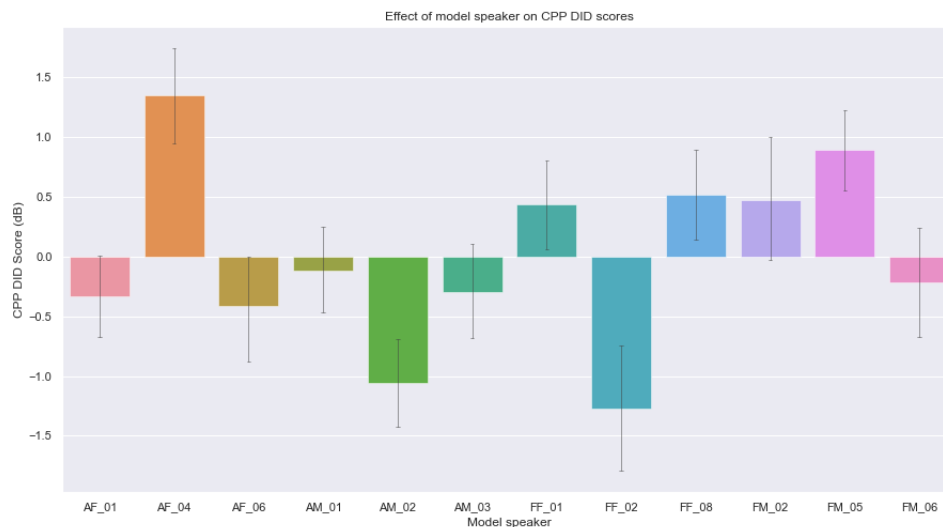


Figure 87: Effect of model speaker on CPP DID scores. Error bars represent the standard deviation of the mean.

These results suggest that there was no significant overall accommodation on CPP. Language, the interaction between Language and Gender, and Model speaker had an effect on CPP DID scores. Overall, subjects converged towards French model speakers and diverged from American English ones. American English female model speakers received convergence but French female model speakers received divergence. The opposite pattern was observed for male speaker. Significant gender-based differences were only observed in American English, with subjects diverging more from male and converging towards female model speakers. Male French model speakers received the most convergence on CPP. Subjects converged more towards AF_04 on CPP and diverged more from FF_02.

2.5 Analysis of EGG measures

2.5.1 Closed quotient DID score

Mean CQ DID scores across language and gender are displayed in Figure 88. As for CPP, the pattern is completely reversed across language. French female learners have overall converged towards male American English model speakers but diverged from French ones, and converged

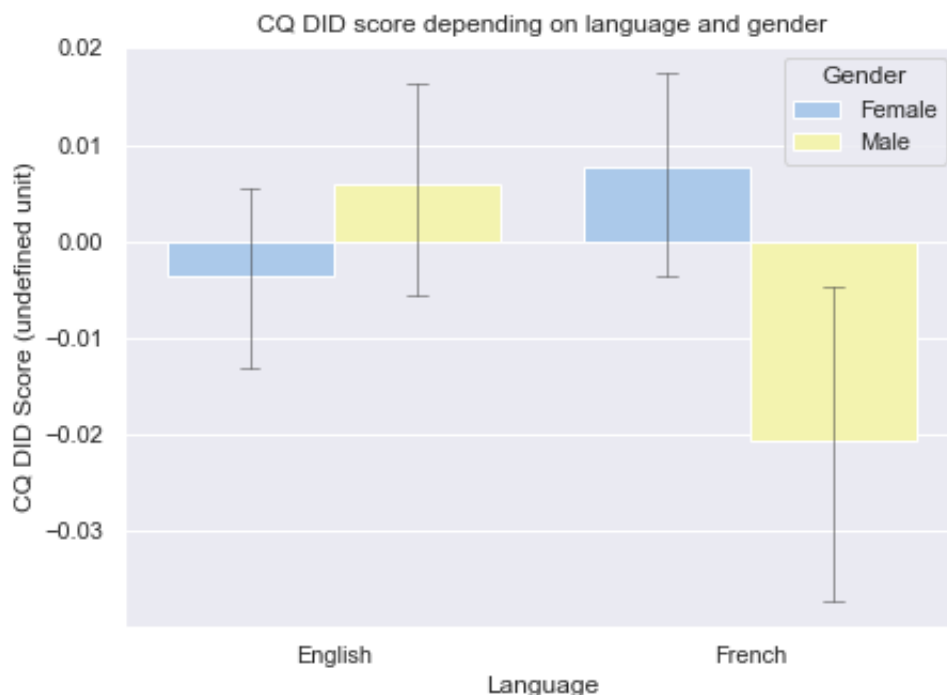


Figure 88: Bar plot representing the mean CQ DID score by language and gender. Error bars represent the standard error of the mean.

towards French female model speakers but diverged from American English ones. Error bars indicate that some subjects have also converged towards American English female model speakers and diverged from both male American English model speakers and French female model speakers.

Table 39: Mean CQ difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively).

| Mean CQ difference (undefined unit) | | |
|-------------------------------------|----------------|------------------|
| Language | <i>French</i> | \searrow 0.006 |
| | <i>English</i> | \nearrow 0.01 |
| Language:Gender | <i>FM</i> | \searrow 0.03 |
| | <i>FF</i> | \nearrow 0.01 |
| | <i>AM</i> | \nearrow 0.01 |
| | <i>AF</i> | \nearrow 0.02 |

Table 39 shows the mean CQ difference values between the reading task and the repetition

tasks according to language and model speakers. Subjects produced slightly lower CQ values when repeating French model speakers and higher CQ values when repeating American English model speakers. If we break down the data across gender, we can observe that subjects have produced higher CQ values when repeating French female model speakers, as well as both female and male American English ones. They only produced decreased CQ values when repeating male French model speakers. Comparing now these values to the CQ DID values displayed in Figure 88, we can conclude that, for convergence effects to have been observed towards female American English, more subjects should have produced lower CQ values during the repetition task. The reverse pattern is observed for convergence towards male French model speakers: subjects should have produced higher CQ values when repeating that group.

To determine if Language and the interaction between Language and Gender had a significant effect on CQ DID scores, we first fitted our control model (as described previously) with CQ DID as the dependent measure. The control MEM's intercept was non-significant ($\beta = 0.003$, $t = -0.31$, $p = .76$), indicating no significant overall accommodation to CQ. Language ($\beta = -7.81e-03$, $t = -1.35$, $p = .18$; $\chi^2(1) = 1.82$, $p = .18$) did not improve model fit, but the interaction between the two ($\beta = -3.77e-02$, $t = -3.26$, $p < .01$; $\chi^2(5) = 14.78$, $p < .01$) and Model speakers ($\chi^2(11) = 4.55$, $p < .05$) did.

Effect of the interaction between language and gender is displayed in Figure 89. *Post-hoc* comparisons showed that differences in CQ DID scores were significant for French male model speakers as compared to French female ($\beta = 0.028$, $t = 3.41$, $p < .01$) and male American English model speakers ($\beta = 0.026$, $t = 3.23$, $p < .01$).

As can be seen in Figure 90, subjects converged more towards FF_08 ($M = 0.02 \pm 0.16$) on CQ, then towards FF_02 ($M = 0.014 \pm 0.1$), AM_03 ($M = 0.011 \pm 0.14$), AM_01 ($M = 0.01 \pm 0.12$), and AF_06 ($M = 0.001 \pm 0.1$). Subjects diverged from all other model speakers, with FM_05 ($M = -0.04 \pm 0.2$) receiving the most divergence on CQ, followed by FM_02 ($M = -0.016 \pm 0.19$), FF_01 ($M = -0.009 \pm 0.1$), FM_06 ($M = -0.007 \pm 0.17$), FF_01 ($M = -0.007 \pm 0.09$), AF_04 ($M = -0.004 \pm 0.13$), and AM_02 ($M = -0.002 \pm 0.14$).

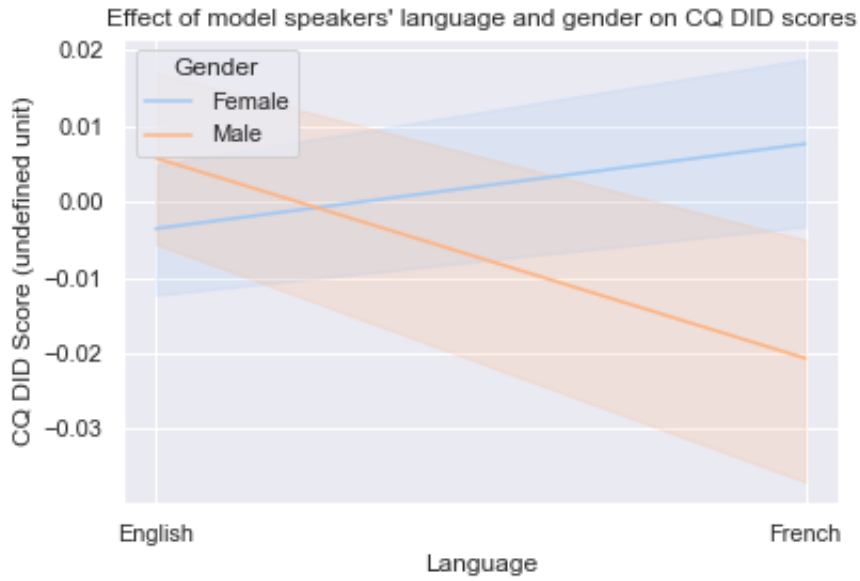


Figure 89: Effect of language and gender on CQ DID scores. The shaded areas represent the standard deviation.

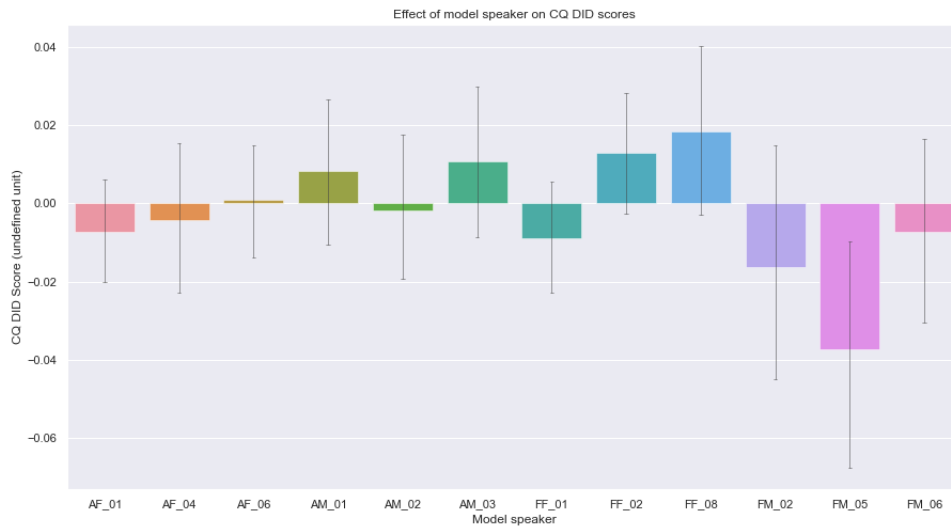


Figure 90: Effect of model speaker on CQ DID scores. Error bars represent the standard deviation of the mean.

These results suggest that there was no significant overall accommodation on CQ. Alignment on CQ did not vary by language. The interaction between Language and Gender, as well as Model speaker, resulted in significant effects. Subjects very slightly converged more towards male

American English model speakers, and diverged from their female counterparts. The reverse pattern was observed in French. Significant gender-based differences were only observed in French, with subjects diverging more from male and converging more towards female model speakers. Female French model speakers received the most convergence on CQ. Subjects converged more towards FF_08 on CQ and diverged more from FM_05.

2.5.2 PIC DID score

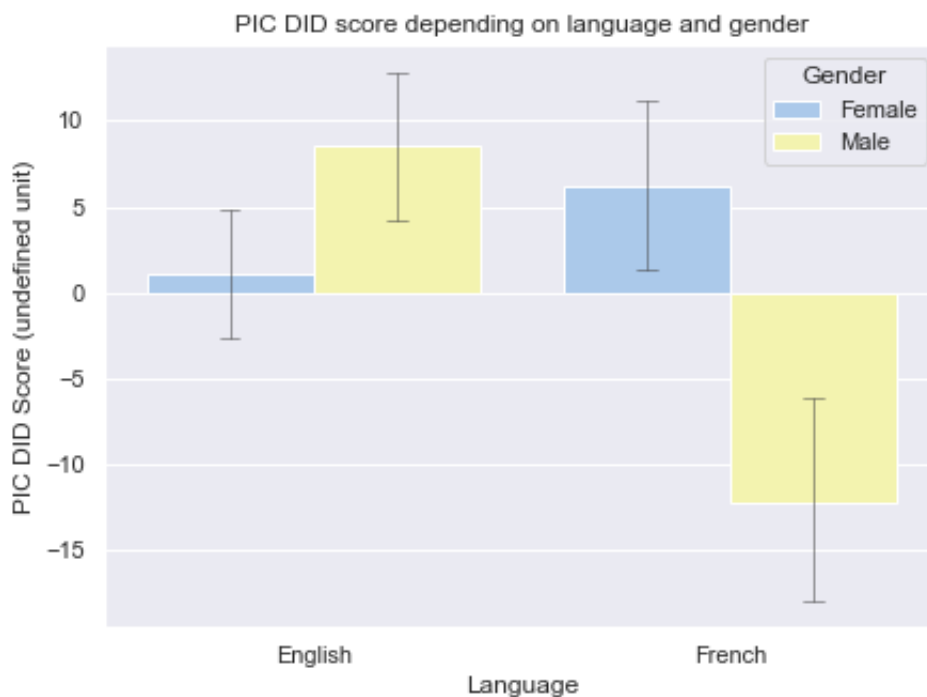


Figure 91: Bar plot representing the mean PIC DID score by language and gender. Error bars represent the standard error of the mean.

Mean PIC DID scores across language and gender are displayed in Figure 91. There is a general trend towards convergence, with a stronger effect towards American English model speakers. Subjects converged overall more towards male American English model speakers, and more towards female French model speakers than towards American English ones. Divergence effects are only observed for male French model speakers. Error bars indicate that some subjects have also diverged from female American English model speakers.

Table 40: Mean PIC difference between the reading task and the repetition task. Increased or decreased mean difference between the two tasks are indicated with the corresponding arrow (\searrow or \nearrow , respectively).

| Mean PIC difference (undefined unit) | | |
|--------------------------------------|----------------|------------------|
| Language | <i>French</i> | \searrow 29.44 |
| | <i>English</i> | \searrow 16.86 |
| Language:Gender | <i>FM</i> | \searrow 43.08 |
| | <i>FF</i> | \searrow 15.81 |
| | <i>AM</i> | \searrow 16.67 |
| | <i>AF</i> | \searrow 16.97 |

Table 40 shows the mean PIC difference values between the reading task and the repetition tasks according to language and model speakers. Subjects produced lower PIC values when repeating all groups. If we now compare these values to the PIC DID values displayed in Figure 91, we can conclude that, for overall convergence effects to have been observed towards male French speakers, subjects should have produced higher PIC values when repeating that group.

To determine if Language and the interaction between Language and Gender had a significant effect on PIC DID scores, we first fitted a control model (as described previously) with PIC DID as the dependent measure. The control MEM's intercept was non-significant ($\beta = 0.86$, $t = 0.18$, $p = .86$), indicating no significant overall accommodation to PIC. Language ($\beta = -7.81$, $t = -3.28$, $p < .01$; $\chi^2(1) = 10.7$, $p < .01$), the interaction between Language and Gender ($\beta = -25.71$, $t = -5.43$, $p < .001$; $\chi^2(5) = 44.53$, $p < .001$), and Model speaker ($\chi^2(11) = 26.74$, $p < .001$) improved model fit.

We can observed in Figure 92 that, overall, subjects significantly converged more towards American English model speakers than towards French model speakers to which they diverged. *Post-hoc* comparisons showed that there was a significant difference in PIC DID scores for French male model speakers as compared to American English male ($\beta = 20.58$, $t = 6.13$, $p < .001$), American English female ($\beta = 13.16$, $t = 5.38$, $p < .001$), and French female model speakers ($\beta = 18.28$, $t = 5.38$, $p < .001$), as illustrated in Figure 93.

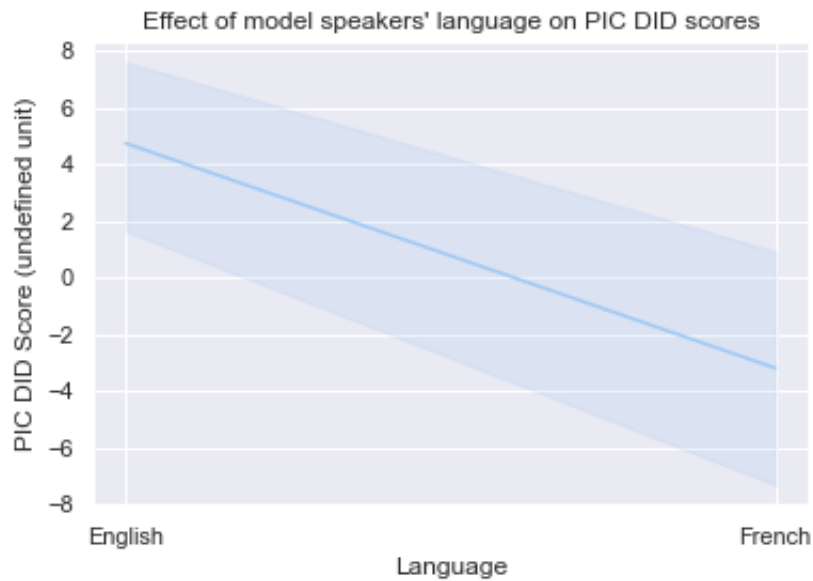


Figure 92: Effect of language on PIC DID scores. The shaded area represents the standard deviation.

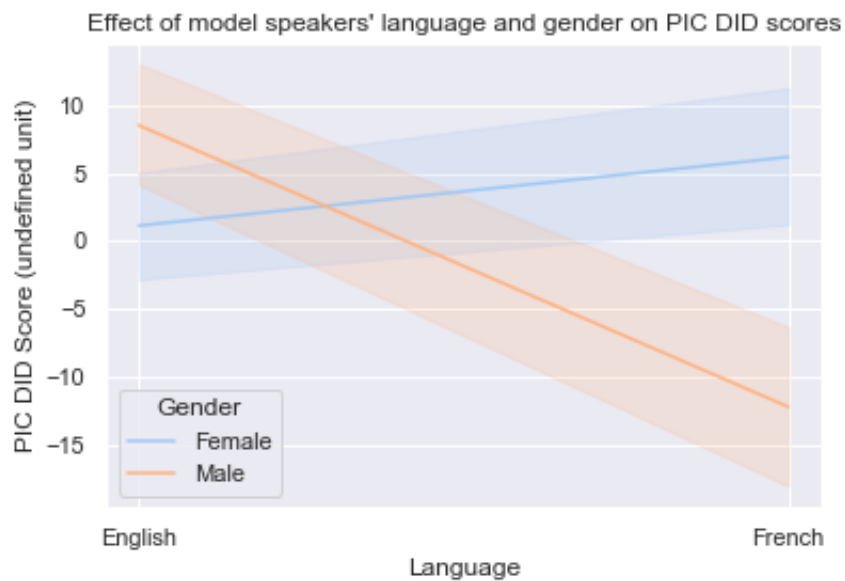


Figure 93: Effect of language and gender on PIC DID scores. The shaded areas represent the standard deviation.

As can be seen in Figure 94, AM_01 ($M = 13.81 \pm 56.21$) received the most convergence on PIC, followed by AM_02 ($M = 9.4 \pm 59.63$), FF_01 ($M = 9.23 \pm 57.24$), AF_04 ($M = 8.29 \pm 55.84$), FF_02 ($M = 6.92 \pm 62.28$), AM_03 ($M = 3.02 \pm 51.7$), and FF_08 ($M = 2.87$

± 60.48). Subjects diverged from all other model speakers, with FM_06 ($M = -16.71 \pm 69.01$) receiving the most divergence, followed by FM_05 ($M = -9.92 \pm 65.49$), FM_02 ($M = -9.14 \pm 80.08$), AF_06 ($M = -3.8 \pm 40.97$), and AF_01 ($M = -1.12 \pm 47.69$).

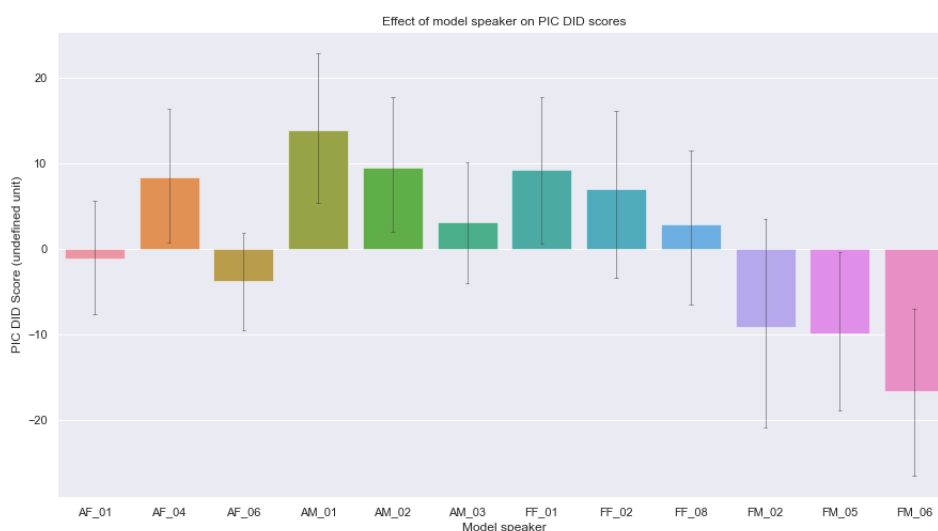


Figure 94: Effect of model speaker on PIC DID scores. Error bars represent the standard deviation of the mean.

These results suggest that there was no significant overall accommodation on PIC. Language, the interaction between Language and Gender, and Model speaker had a effect on PIC DID scores. Overall, subjects converged towards American English model speakers and diverged from French ones. Both male and female American English model speakers received convergence. Female French model speakers received convergence and male French model speakers divergence. Significant gender-based differences were only observed in French, with subjects diverging from male and converging towards female model speakers. Male American English model speakers received the most alignment on PIC. Regarding Model speaker, AM_01 received the most convergence on PIC while FM_06 received the most divergence.

In this section we analysed convergence effects on various measurements often used in phonation studies. Our goal was to observe whether French learners of English managed to converge towards model speakers, and, therefore, successfully imitated creaky voice. A summary of the

accommodation patterns for acoustic and EGG measures is given in Table 41 and 42, respectively⁶². We cannot conclude that French learners of English overall successfully imitated creaky voice. There was a language effect, with subjects significantly converging on more dimensions towards American English than towards French model speakers. Significant divergence were observed on more dimensions from French model speakers. Dimensions on which subjects converged or diverged differed across language. We observed significant convergence on more dimensions towards male than towards female French model speakers, and significant divergence on more dimensions from female than from male French model speakers. Subjects converged significantly on the same number of dimensions towards either male or female American English model speakers, but diverged significantly more from male than from female American English model speakers. Model speaker seems to have accounted for more variation in accommodation. A great deal of variability was observed in convergence/divergence effects depending on Model speaker and dimensions, indicating that listeners might converge more towards specific speakers⁶³. Our results also confirm that subjects can converge on dimensions while diverging on others (Pardo, 2012).

In the next section we analyse results of the rating task we conducted to determine whether creaky voice was overall perceived more positively than non-creaky voice, and whether language and the interaction between Language and Gender had an effect on the evaluation of creaky voice.

⁶²As evidenced by the R^2 (see Appendix B.2), most of our models actually captured very little of the variance observed.

⁶³A summary of the convergence and divergence effects for each measure depending on model speakers is given in Appendix B.1.

Table 41: Synthesis of the significant patterns of accommodation observed on acoustic measures in our dataset. Overall significant patterns are reported in the left column, those across language in the middle column, those resulting from the interaction between Language and Gender in the right column. An asterisk has been added whenever the pattern was the same across groups to indicate what group received the most convergence/divergence. Blank spaces mean no significant patterns of accommodation.

| Acoustic measures | Overall pattern of accommodation | Language | Language:Gender | |
|----------------------|----------------------------------|----------------|---------------------------------------|-----------------------------|
| Duration | Convergence | <i>French</i> | <i>FM</i> <i>FF</i> | Maintenance Convergence* |
| | | <i>English</i> | <i>AM</i> <i>AF</i> | Convergence Convergence |
| <i>f₀</i> | Divergence | <i>French</i> | <i>FM</i> <i>FF</i> | Convergence Divergence |
| | | <i>English</i> | Convergence <i>AM</i> <i>AF</i> | Convergence* Maintenance |
| H1* | Divergence | <i>French</i> | Divergence* <i>FM</i> <i>FF</i> | Divergence* Divergence |
| | | <i>English</i> | Divergence <i>AM</i> <i>AF</i> | Divergence Divergence |
| H2* | Divergence | <i>French</i> | Divergence* <i>FM</i> <i>FF</i> | Divergence* Divergence |
| | | <i>English</i> | Divergence <i>AM</i> <i>AF</i> | Divergence Divergence |
| H1*-H2* | | <i>French</i> | Divergence <i>FM</i> <i>FF</i> | Divergence |
| | | <i>English</i> | Maintenance <i>AM</i> <i>AF</i> | Maintenance |
| H1*-A1* | Divergence | <i>French</i> | <i>FM</i> <i>FF</i> | Divergence |
| | | <i>English</i> | <i>AM</i> <i>AF</i> | Divergence |
| H1*-A2* | Divergence | <i>French</i> | <i>FM</i> <i>FF</i> | Divergence Divergence |
| | | <i>English</i> | <i>AM</i> <i>AF</i> | Divergence* Maintenance |
| H1*-A3* | Divergence | <i>French</i> | Divergence* <i>FM</i> <i>FF</i> | Divergence* Divergence |
| | | <i>English</i> | Divergence <i>AM</i> <i>AF</i> | Divergence Convergence |
| H2*-H4* | | <i>French</i> | <i>FM</i> <i>FF</i> | Divergence |
| | | <i>English</i> | <i>AM</i> <i>AF</i> | Divergence |
| CPP | Convergence | <i>French</i> | <i>FM</i> <i>FF</i> | Convergence* Divergence |
| | | <i>English</i> | Divergence <i>AM</i> <i>AF</i> | Divergence* Convergence |

Table 42: Synthesis of the significant patterns of accommodation observed on EGG measures in our dataset. Overall significant patterns are reported in the left column, those across language in the middle column, those resulting from the interaction between Language and Gender in the right column. An asterisk has been added whenever the pattern was the same across groups to indicate what group received the most convergence/divergence. Blank spaces mean no significant patterns of accommodation.

| EGG measures | Overall pattern of accommodation | Language | Language:Gender | |
|--------------|----------------------------------|----------------|-----------------|--------------|
| CQ | | <i>French</i> | <i>FM</i> | Divergence* |
| | | | <i>FF</i> | Convergence* |
| | | <i>English</i> | <i>AM</i> | Convergence |
| | | | <i>AF</i> | Divergence |
| PIC | | <i>French</i> | <i>FM</i> | Divergence |
| | | | <i>FF</i> | Convergence |
| | | <i>English</i> | <i>AM</i> | Convergence* |
| | | | <i>AF</i> | Convergence |

3 Rating task

The experiment followed a 'matched guise' design in which subjects were presented with some of the same stimuli used in the reading and imitation task. They were asked to evaluate their impression of some model speaker's voice on 4 six point-point semantic scales with anchors of 'not at all' (1) and 'very' (6). We selected 3 same sentences produced in both qualities (non-creaky/creaky) by 8 model speakers (2 American male, 2 American female, 2 French male, 2 French female)⁶⁴. Our goal was to evaluate the attitude towards creaky voice and observed whether there were any between-language effects and gender-based differences observed within-language.

The influence of our variables on voice quality evaluation are first reported graphically. We then report the results of the different ordinal logistic regression models we built using the function `polr()` of the *MASS* package in RStudio 2022.7.1 (Venables & Ripley, 2002). Separate models were created for each trait. They respect the following pattern.

```
modtrait <- polr(Trait ~ VQ + VQ:Language + VQ:Language:Gender)
```

Overall observations are provided in section 3.1. The effect of Language on voice quality evaluation is analysed in section 3.2 and Gender effect in section 3.3. Finally, we analyse Model speaker effect in section 3.4.

3.1 Overall observations

Figure 95 shows the averaged ratings of model speakers' voice when producing non-creaky and creaky voice in word-final position. Ratings of creaky voice are displayed in pink and that of non-creaky voice in green. The different features are indicated on the x-axis and the averaged rating on the y-axis⁶⁵. There does not seem to be significant differences between how the two voice qualities were evaluated. However, overall higher ratings were observed for non-creaky voice. Creaky voice was only evaluated as sounding more attractive than non-creaky voice. All

⁶⁴The reader should refer to Chapter III, section 4.2.3 for a more detailed description of the task.

⁶⁵Subsequent graphs will follow the same pattern.

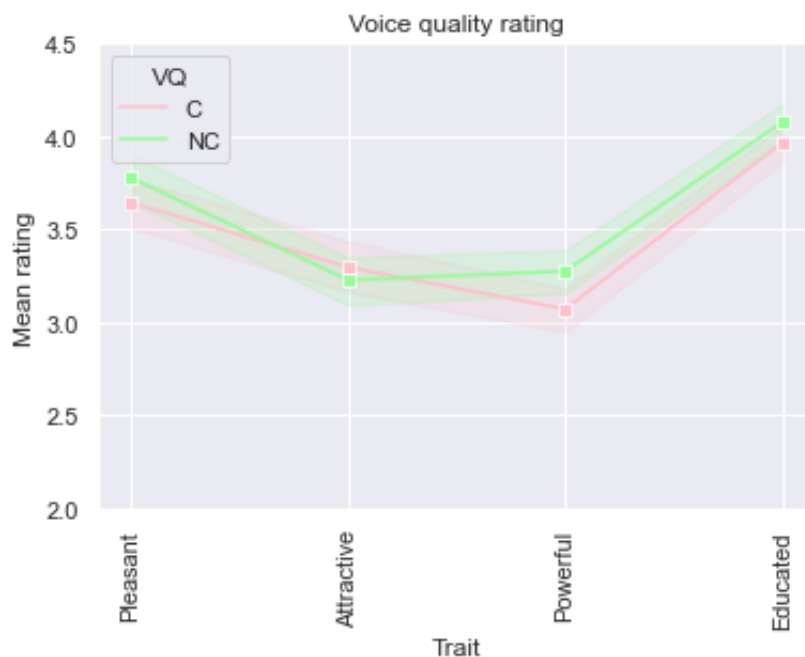


Figure 95: Voice quality rating across the whole dataset. Creaky voice is represented by the pink line, non-creaky voice by the green one. Shaded areas represent the standard deviation of the mean.

model speakers' voices, when producing both creaky or non-creaky voice in word-final position, were evaluated more positively than negatively (mean above 3). From more positively rated to less positively rated, model speakers' voices sounded more educated ($M_C = 3.97 \pm 1.11$; $M_{NC} = 4.08 \pm 1.09$), pleasant ($M_C = 3.64 \pm 1.49$; $M_{NC} = 3.78 \pm 1.41$), attractive ($M_C = 3.3 \pm 1.56$; $M_{NC} = 3.23 \pm 1.52$), and powerful ($M_C = 3.07 \pm 1.31$; $M_{NC} = 3.31 \pm 1.3$). This pattern is the same across voice qualities.

The interpretation of the coefficients resulting from our models showed that creaky voice in word-final position, as opposed to non-creaky voice, is associated with:

- no likelihood of being perceived as less pleasant
- no likelihood of being perceived as less attractive
- no likelihood of being perceived as less powerful
- no likelihood of being perceived as less educated

The odds of creaky voice in word-final position to be perceived as less pleasant are only 1.15

times that of non-creaky voice (95% CI: [0.681, 1.962])⁶⁶, as more attractive 1 time that of non-creaky voice [0.589, 1.69], as less powerful 1.25 times that of non-creaky voice [0.73, 2.144], and as less educated 0.78 times that of non-creaky voice [0.439, 1.378], holding constant all variables.

3.2 Language effect

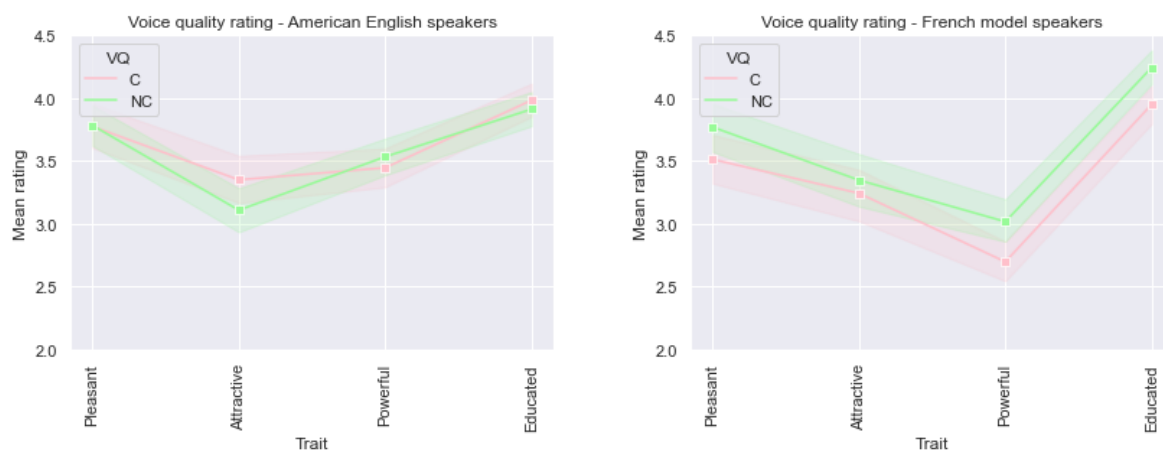


Figure 96: Voice quality rating across language. Ratings of American English voices are displayed on the left panel, those of French on the right panel. Creaky voice is represented by the pink line, non-creaky voice by the green one. Shaded areas represent the standard deviation of the mean.

The general tendency previously observed might not be reflected across language. Figure 96 shows the averaged rating of model speakers' voice when producing non-creaky and creaky voice in word-final position in both languages. Ratings of American English model speakers' voices are displayed on the left panel, those of French speakers' on the right panel.

3.2.1 American English

As observed in Figure 96, there does not seem to be significant differences between how the two voice qualities were evaluated in American English. However, overall higher ratings were observed for creaky voice, although the differences are very subtle. Creaky voice was considered more attractive and to sound slightly more educated. Both creaky and non-creaky voice sounded

⁶⁶All reported confidence intervals are 95%.

similarly pleasant, and non-creaky voice was rated a slightly more powerful. All American English models' voices, when producing both creaky or non-creaky voice, were evaluated more positively than negatively (mean above 3). From more positively rated to less positively rated, model speakers' voices sounded more educated ($M_C = 3.98 \pm 1.02$; $M_{NC} = 3.91 \pm 1.06$), pleasant ($M_C = 3.78 \pm 1.38$; $M_{NC} = 3.78 \pm 1.34$), powerful ($M_C = 3.45 \pm 1.2$; $M_{NC} = 3.53 \pm 1.22$), and attractive ($M_C = 3.35 \pm 1.46$; $M_{NC} = 3.11 \pm 1.42$). This pattern is the same across voice qualities.

The interpretation of the coefficients resulting from our models showed that American creaky voice, as opposed to American non-creaky voice, is associated with:

- no likelihood of being perceived as less pleasant
- no likelihood of being perceived as more attractive
- no likelihood of being perceived as less powerful
- no likelihood of being perceived as more educated

The odds of American creaky voice to be perceived as less pleasant are only 1.16 times that of creaky voice [0.679, 1.999], as more attractive are 0.99 times that of non-creaky voice [0.85, 2.637], as less powerful are 1.26 times that of non-creaky voice [0.733, 2.175], as more educated are 0.77 times that of American non-creaky voice [0.432, 1.383], holding constant all variables.

3.2.2 French

Again, Figure 96 shows no significant differences between how the two voice qualities were evaluated in French. However, overall higher ratings were observed for non-creaky voice. All French models' voices were evaluated more positively than negatively when producing non-creaky voice (mean above 3) in word-final position. Creaky voice was overall more positively rated than negatively, except for 'powerful' which received a negative rating (mean less than 3), suggesting it does not sound really powerful. From more positively rated to less positively rated, model speakers' voices sounded more educated ($M_C = 3.95 \pm 1.19$; $M_{NC} = 4.24 \pm 1.1$), pleasant (M_C

= 3.51 ± 1.59 ; $M_{NC} = 3.77 \pm 1.46$), attractive ($M_C = 3.24 \pm 1.66$; $M_{NC} = 3.35 \pm 1.61$), and powerful ($M_C = 2.7 \pm 1.31$; $M_{NC} = 3.02 \pm 1.33$). This pattern is the same across voice qualities.

The interpretation of the coefficients resulting from our models showed that French creaky voice in word-final position, as opposed to French non-creaky voice, is associated with:

- no likelihood of being perceived as less pleasant
- no likelihood of being perceived as less attractive
- small likelihood of being perceived as less powerful ($p < .01$)
- small likelihood of being perceived as less educated ($p < .05$)

The interpretation of the odds ratio resulting from our models showed that the odds of French creaky voice in word-final position to be perceived as less pleasant are only 1.78 times that of creaky voice [0.975, 3.078], as less attractive 1.5 times that of creaky voice [0.85, 2.637], as less powerful are 2.38 times that of creaky voice [1.353, 4.226], as less educated are 1.81 times that of French non-creaky voice [1.025, 3.218], holding constant all variables.

3.2.3 American English vs. French

American English model speakers' voices were overall more positively rated than French ones, whenever models would use creaky or non-creaky voice. Only non-creaky American English voices were rated as slightly less educated than French ones. Mean ratings for all traits were higher for creaky voice in American English as compared to French. Creaky voice is, therefore, more positively perceived in American English.

The interpretation of the coefficients resulting from our models showed that when creaky voice is produced by French speakers, as opposed to when produced by American English speakers, it is associated with:

- no likelihood of being perceived as less pleasant
- no likelihood of being perceived as less attractive
- no likelihood of being perceived as less powerful

- no likelihood of being perceived as less educated

The interpretation of the odds ratio resulting from our models showed that the odds of creaky voice when produced by French model speakers to be perceived as less pleasant are only 1.12 times that of when produced by American English speakers [0.64, 1.962], as less attractive 1.43 times that of when produced by American English speakers [0.823, 2.496], as less powerful 0.62 times that of when produced by American English speakers [0.353, 1.084], and as less educated 1.54 times that of when produced by American English speakers [0.871, 2.746], holding constant all variables.

Although our models did not result in significant effects, creaky voice seems to be more positively perceived in American English than in French. In the next section we look at whether gender affected voice quality rating within-language.

3.3 Gender effect within-language

Figure 97 shows the averaged ratings of male and female model speakers' voices when producing non-creaky and creaky voice in word-final position in both languages. Ratings of male American English model speakers' voices are displayed on top left panel, those of female American English on the top right panel. Ratings of male French model speakers' voices are displayed on the bottom left panel, those of female French on the bottom right panel.

3.3.1 American English male

We can observe in Figure 97 some differences between how the two voice qualities were evaluated when produced by male American English speakers. All male American English voices were overall evaluated more positively than negatively (mean above 3). Only one trait received a negative judgment: non-creaky voice was judged as being less attractive when produced by male American English speakers. Creaky voice was perceived as more pleasant, more attractive, and to sound more educated but less powerful than non-creaky voice. From more positively rated to less positively rated, male American English model speakers' voices sounded more educated

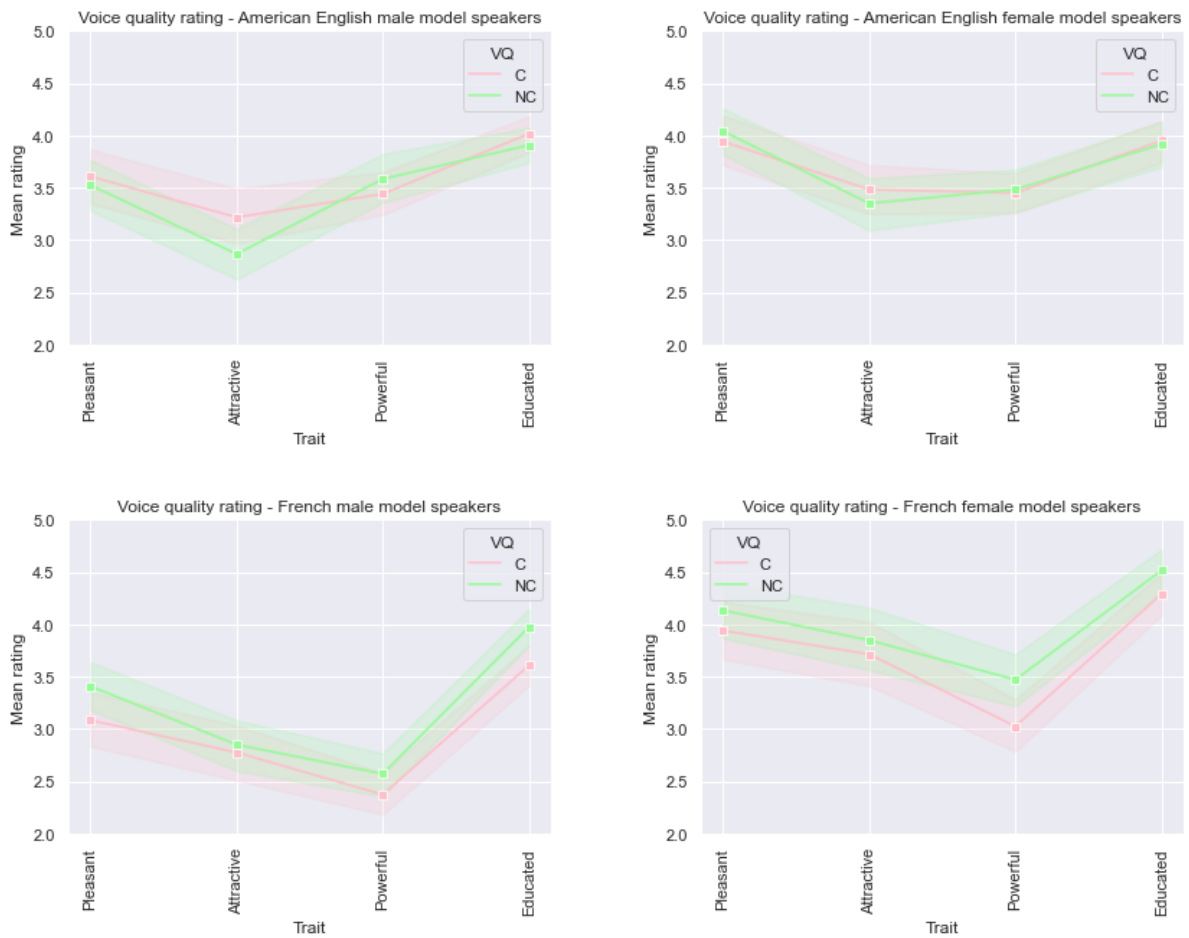


Figure 97: Voice quality rating across language and gender. Ratings of American English voices are displayed on the left panel, those of French on the right. Ratings of male voices are displayed on the top panels, those of female on the bottom ones. Creaky voice is represented by the pink line, non-creaky voice by the green one. Shaded areas represent the standard deviation of the mean.

($M_C = 4.02 \pm 0.95$; $M_{NC} = 3.91 \pm 0.94$), pleasant ($M_C = 3.61 \pm 1.4$; $M_{NC} = 3.53 \pm 1.4$), powerful ($M_C = 3.44 \pm 1.26$; $M_{NC} = 3.58 \pm 1.28$), and attractive ($M_C = 3.22 \pm 1.49$; $M_{NC} = 2.87 \pm 1.41$). This pattern is the same across voice qualities.

The interpretation of the coefficients resulting from our models showed that male American English creaky voice in word-final position, as opposed to male American English non-creaky voice, is associated with:

- no likelihood of being perceived as more pleasant
- small likelihood of being perceived as more attractive ($p < .05$)

- no likelihood of being perceived as less powerful
- no likelihood of being perceived as more educated

The odds of creaky voice in word-final position when produced by male American English speakers to be perceived as more pleasant are only 0.78 times that of non-creaky voice [0.453, 1.353], as more attractive are 0.56 times that of non-creaky voice [0.321, 0.967], as less powerful are 0.95 times that of non-creaky voice [0.552, 1.653], as more educated are 0.67 times that of non-creaky voice [0.38, 1.188], holding constant all variables.

3.3.2 American English female

There is close to no differences between how the two voice qualities were evaluated when produced by female American English speakers, as shown in Figure 97. All female American English voices were overall evaluated more positively than negatively (mean above 3). Creaky voice was only perceived as being slightly more attractive than non-creaky voice. From more positively rated to less positively rated, female American English model speakers' voices sounded more pleasant ($M_C = 3.94 \pm 1.34$; $M_{NC} = 4.04 \pm 1.29$), educated ($M_C = 3.95 \pm 1.09$; $M_{NC} = 3.92 \pm 1.18$), powerful ($M_C = 3.45 \pm 1.14$; $M_{NC} = 3.48 \pm 1.17$), and attractive ($M_C = 3.48 \pm 1.43$; $M_{NC} = 3.35 \pm 1.39$). This pattern is the same across voice qualities.

The interpretation of the coefficients resulting from our models showed that female American English creaky voice in word-final position, as opposed to female American English non-creaky voice, is associated with:

- no likelihood of being perceived as less pleasant
- no likelihood of being perceived as more attractive
- no likelihood of being perceived as less powerful
- no likelihood of being perceived as more educated

The odds of creaky voice in word-final position when produced by female American English speakers to be perceived as less pleasant are only 1.17 times that of non-creaky voice [0.676,

2.022], as more attractive are 0.98 times that of non-creaky voice [0.568, 1.693], as less powerful are 1.27 times that of non-creaky voice [0.736, 2.222], as more educated are 0.8 times that of non-creaky voice [0.459, 1.386], holding constant all variables.

3.3.3 French male

As observed in Figure 97, there does not seem to be significant differences between how the two voice qualities were evaluated when produced by male French speakers. Depending on traits, male French voices were either more positively evaluated or more negatively (mean above or below 3). Voices were, for instance, perceived as being not very attractive or powerful when produced with either word-final creak or without word-final creak. Creaky voice was always perceived more negatively than non-creaky voice. From more positively rated to less positively rated, male French model speakers' voices sounded more educated ($M_C = 3.61 \pm 1.07$; $M_{NC} = 3.98 \pm 0.99$) and pleasant ($M_C = 3.08 \pm 1.37$; $M_{NC} = 3.41 \pm 1.28$), but not so attractive ($M_C = 2.78 \pm 1.45$; $M_{NC} = 2.85 \pm 1.35$) and powerful ($M_C = 2.37 \pm 1.14$; $M_{NC} = 2.57 \pm 1.12$). This pattern is the same across voice qualities.

The interpretation of the coefficients resulting from our models showed that male French creaky voice, as opposed to male French non-creaky voice, is associated with:

- no likelihood of being perceived as less pleasant
- no likelihood of being perceived as less attractive
- no likelihood of being perceived as less powerful
- no likelihood of being perceived as less educated

The odds of creaky voice when produced by male French speakers to be perceived as less pleasant are only 1.67 times that of non-creaky voice [0.966, 2.918], as less attractive are 1.08 times that of non-creaky voice [0.625, 1.87], as less powerful are 1.23 times that of non-creaky voice [0.704, 2.138], as less educated are 1.62 times that of non-creaky voice [0.924, 2.862], holding constant all variables.

3.3.4 French female

There does not seem to be significant differences between how the two voice qualities were evaluated when produced by female French speakers, as shown in Figure 97. All female French voices were overall evaluated more positively than negatively (mean above 3). Creaky voice was always perceived more negatively than non-creaky voice. From more positively rated to less positively rated, male French model speakers' voices sounded more educated ($M_C = 4.29 \pm 1.21$; $M_{NC} = 4.52 \pm 1.14$), pleasant ($M_C = 3.94 \pm 1.68$; $M_{NC} = 4.13 \pm 1.53$), attractive ($M_C = 3.71 \pm 1.72$; $M_{NC} = 3.85 \pm 1.7$) and powerful ($M_C = 3.03 \pm 1.39$; $M_{NC} = 3.47 \pm 1.36$). This pattern is the same across voice qualities.

The interpretation of the coefficients resulting from our models showed that female French creaky voice, as opposed to female French non-creaky voice, is associated with:

- no likelihood of being perceived as less pleasant
- no likelihood of being perceived as less attractive
- small likelihood of being perceived as less powerful ($p < .01$)
- no likelihood of being perceived as less educated

The odds of creaky voice when produced by female French speakers to be perceived as less pleasant are only 1.59 times that of non-creaky voice [0.918, 2.775], as less attractive are 1.42 times that of non-creaky voice [0.823, 2.455], as less powerful are 2.22 times that of non-creaky voice [1.274, 3.91], as less educated are 1.72 times that of non-creaky voice [0.984, 3.026], holding constant all variables.

3.3.5 Male vs. Female American English

We saw in the previous section that, depending on traits, creaky or non-creaky voice in word-final position was more positively rated within-gender in American English. If we compare the mean ratings that each trait got across gender in American English, we can conclude that female model speakers' voices were overall more positively rated than male ones, whenever models would

use creaky or non-creaky voice in word-final position. Creaky voice is, therefore, overall more positively perceived when produced by female than by male American English speakers.

The interpretation of the coefficients resulting from our models showed that, when creaky voice is produced by male American English speakers, as opposed to when produced by female American English speakers, it is associated with:

- small likelihood of being perceived as less pleasant ($p < .05$)
- no likelihood of being perceived as less attractive
- no likelihood of being perceived as more powerful
- no likelihood of being perceived as less educated

The odds of creaky voice when produced by male American English speakers to be perceived as less pleasant are only 0.55 times that of when produced by female American English speakers [0.316, 0.943], as less attractive are 0.66 times that of when produced by female American English speakers [0.378, 1.141], as more powerful are 1 time that of when produced by female American English speakers [0.581, 1.731], as less educated are 0.96 than by when produced by female American English speakers [0.547, 1.669], holding constant all variables.

3.3.6 Male vs. Female French

We saw in the previous section that creaky voice was more negatively perceived within-gender in French. If we compare the mean ratings that each trait got across gender in French, we can conclude that female model speakers' voices were overall more positively rated than male ones, whenever models would use creaky or non-creaky voice. As for American English speakers, creaky voice is, therefore, overall more positively perceived when produced by female than by male French speakers, although it is still perceived more negatively than non-creaky voice.

The interpretation of the coefficients resulting from our models showed that creaky voice, when produced by male French speakers, as opposed to when produced by female French speakers, is associated with:

- small likelihood of being perceived as less pleasant ($p < .01$)

- small likelihood of being perceived as less attractive ($p < .01$)
- small likelihood of being perceived as less powerful ($p < .05$)
- small likelihood of being perceived as less educated ($p < .01$)

The odds of creaky voice, when produced by male French speakers, to be perceived as less pleasant are 0.49 times that of non-creaky voice as opposed to when produced by female French speakers [0.276, 0.859], as less attractive are 0.43 times that of non-creaky voice as opposed to when produced by female French speakers [0.243, 0.747], as less powerful are 0.5 times that of non-creaky voice as opposed to when produced by female French speakers [0.286, 0.886], as less educated are 0.47 times that of non-creaky voice as opposed to when produced by female French speakers [0.264, 0.822], holding constant all variables.

3.4 Model speaker effect

In the analysis of alignment effects in the imitation task, we observed an effect of Model speaker on alignment on almost all measures, indicating that speakers converged/diverged more towards particular speakers than towards particular groups. For this reason, we decided to look at whether some model speakers' voices were more positively judged than others. The averaged ratings of model speakers' voices when producing non-creaky and creaky voice in word-final position is displayed in Figure 98. The different features are indicated on the x-axis and the averaged rating on the y-axis.

It appears that all model speakers, when producing either creaky or non-creaky voice in word-final position, were rated as sounding pretty 'educated'. There was much more variation for all other traits, and especially for 'pleasant' and 'attractive'. It confirms that creaky voice was more positively rated when produced by American English speakers. Ratings were such that FF_08 was deemed to be perceived as having the most pleasant and attractive voice, and to be deemed very educated when producing either creaky or non-creaky voice in word-final position. Conclusions that we can draw is that listeners probably relate to other voice/speech features when judging a voice.



Figure 98: Voice quality rating across model speakers. One colour corresponds to one speaker. The continuous represents creaky voice and the dashed line non-creaky voice. Shaded areas represent the standard deviation of the mean.

We created density plots to observe more precisely how each model speaker’s voice was rated on each trait depending on whether they were producing creaky or non-creaky voice in word-final position. We divide the analysis into four subsections, each corresponding to the four traits we used in this task.

3.4.1 Pleasant

Figure 99 shows whether model speakers’ voices were perceived as being more pleasant when producing creaky or non-creaky voice in word-final position. Ratings of creaky voice are displayed on the left panel and those of non-creaky voice on the right panel. Peaks for each model speaker are less distinct in the creaky voice scenario, indicating more variation in how this voice quality was perceived by subjects. They are more clearly located towards the left for non-creaky voice, indicating that creaky voice is overall perceived as being more pleasant than non-creaky voice. We can see that, for some speakers, there is no clear variation when the same model produced

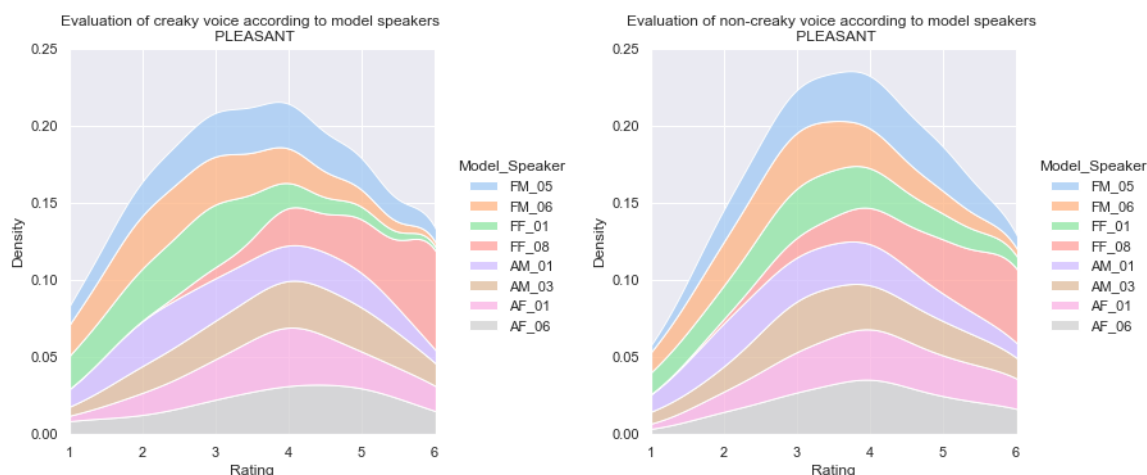


Figure 99: Probability density function of PLEASANT for the different model speakers. Ratings of creaky voice are illustrated on the left panel, those of non-creaky voice on the right panel.

either creaky or non-creaky voice in word-final position (i.e. AF_01, FM_06, and FM_05). Creaky voice seems to be perceived as slightly more pleasant when produced by all other model speakers. If we compare ratings of voice qualities within-speaker, the coefficients resulting from our models showed that creaky voice in word-final position, as opposed to non-creaky voice, is associated with:

- small likelihood of being perceived as less pleasant for FF_01 ($p < .05$)
- no likelihood of being perceived as less pleasant for AF_01, FM_05, and FM_06
- no likelihood of being perceived as more pleasant for AF_06, AM_01, AM_03, and FF_08

The odds of creaky voice in word-final position to be perceived as less pleasant are 1.51 times that of non-creaky voice for AF_01 [0.697, 3.289], 2.5 times that of non-creaky voice for FF_01 [1.14, 5.62], 1.42 that of non-creaky voice for FM_05 [0.659, 3.072], 1.89 that of non-creaky voice for FM_06 [0.871, 4.158]. The odds of creaky voice in word-final position to be perceived as more pleasant are 0.97 times that of non-creaky voice for AF_06 [0.451, 2.089], 0.81 that of non-creaky voice for AM_01 [0.376, 1.744], 0.69 that of non-creaky voice for AM_03 [0.319, 1.486], 0.79 that of non-creaky voice for FF_08 [0.35, 1.782].

If we compare evaluation of creaky voice depending on model speakers, the interpretation of the coefficients resulting from our models showed that, with AF_01 as reference, creaky voice in word-final position is associated with:

- no likelihood of being perceived as more pleasant for AF_06
- no likelihood of being perceived as less pleasant for AM_03 and FM_05
- small likelihood of being perceived as more pleasant for FF_08 ($p < .001$)
- small likelihood of being perceived as less pleasant for AM_01 ($p < .05$), FF_01 ($p < .001$), and FM_06 ($p < .01$)

The odds of creaky voice in word-final position to be perceived as more pleasant are only 1.14 times when produced by AF_06 [0.524, 2.498] and 9.08 times when produced by FF_08 [4.052, 20.799] than when produced by AF_01. The odds of creaky voice in word-final position to be perceived as less pleasant are only 0.45 times when produced by AM_01 [0.208, 0.958], 0.76 times when produced by AM_03 [0.346, 1.665], 0.18 times when produced by FF_01 [0.083, 0.39], 0.55 times when produced by FM_05 [0.256, 1.171], and 0.26 times when produced by FM_06 [0.116, 0.581] than when produced by AF_01.

3.4.2 Attractive

Figure 100 shows whether model speakers' voices were perceived as being more attractive when creaky or non-creaky. Ratings of creaky voice are displayed on the left panel and those of non-creaky voice on the right panel. Peaks for each model speaker are once again less distinct in the creaky voice scenario, indicating more variation towards how this voice quality was perceived by subjects. They seem to be located similarly across voice qualities, indicating not much overall difference in terms of how attractive a model speaker's voice was rated when producing creaky and non-creaky voice in word-final position. We can see that, for some speakers, there is no clear variation when a model produced either creaky or non-creaky voice in word-final position (i.e. AM_01, FF_01, and FM_06, and FF_05). Creaky voice seems to be perceived as slightly more

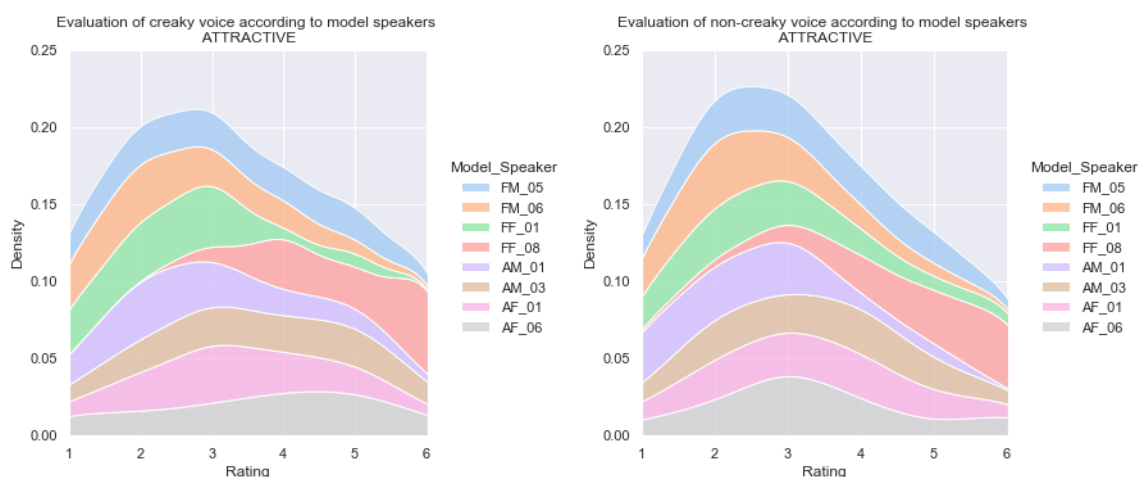


Figure 100: Probability density function of ATTRACTIVE for the different model speakers. Ratings of creaky voice are illustrated on the left panel, those of non-creaky voice on the right panel.

attractive when produced by all other model speakers. If we compare ratings of voice qualities within-speaker, the coefficients resulting from our models showed that creaky voice in word-final position, as opposed to non-creaky voice, is associated with:

- small likelihood of being perceived as more attractive for AM_01 ($p < .05$)
- no likelihood of being perceived as less attractive for AF_01, FF_01, FM_05, and FM_06
- no likelihood of being perceived as more attractive for AF_06, AM_03, and FF_08

The odds of creaky voice in word-final position to be perceived as more attractive are 1.37 times that of non-creaky voice for AF_01 [0.636, 2.96], 1.66 times that of non-creaky voice for FF_01 [0.762, 3.647], 1.02 times that of non-creaky voice for FM_05 [0.477, 2.189], 1.09 for FM_06 [0.503, 2.358], holding all variables constant. The odds of creaky voice in word-final position to be perceived as less attractive are 0.59 times that of non-creaky voice for AF_06 [0.268, 1.269], 0.42 times that of non-creaky voice for AM_01 [0.187, 0.908], 0.57 times that of non-creaky voice for AM_03 [0.262, 1.22], 0.88 times that of non-creaky voice for FF_08 [0.394, 1.954], holding all variables constant.

If we compare evaluation of creaky voice depending on model speakers, the interpretation of

the coefficients resulting from our models, with AF_01 as reference, showed that creaky voice in word-final position is associated with:

- no likelihood of being perceived as more attractive for AF_06, AM_03
- no likelihood of being perceived as less attractive for AM_01 and FM_05
- small likelihood of being perceived as more attractive for FF_08 ($p < .001$)
- small likelihood of being perceived as less attractive for FF_01 ($p < .01$) and FM_06 ($p < .05$)

The odds of creaky voice in word-final position to be perceived as more attractive are only 1.61 times when produced by AF_06 [0.739, 3.522], 1.28 times when produced by AM_03 [0.57, 2.877], 13.41 times when produced by FF_08 [5.96, 30.753] than when produced by AF_01. The odds of creaky voice in word-final position to be perceived as less attractive are only 0.57 times when produced by AM_01 (95% CI: [0.267, 1.227]), 0.3 times when produced by FF_01 [0.138, 0.645], 0.92 times when produced by FM_05 [0.421, 1.988] and 0.44 times when produced by FM06 [0.196, 0.973] than when produced by AF_01.

3.4.3 Powerful

Figure 101 shows whether model speakers' voices were perceived as sounding more powerful when producing creaky or non-creaky voice in word-final position. Ratings of creaky voice are displayed on the left panel and those of non-creaky voice on the right panel. Peaks for each model speaker are once again less distinct in the creaky voice scenario, indicating more variation towards how this voice quality was perceived. They seem to be located similarly across voice qualities, indicating that there is not much overall difference in terms of how powerful a model speaker's voice was rated when producing creaky or non-creaky voice in word-final position. We can see that for some speakers there is no clear difference when a model produced creaky or non-creaky voice in word-final position (i.e. FF_08, FF_01, FM_06, and FF_05). Creaky voice seems to be perceived as sounding slightly more powerful when produced by all other model

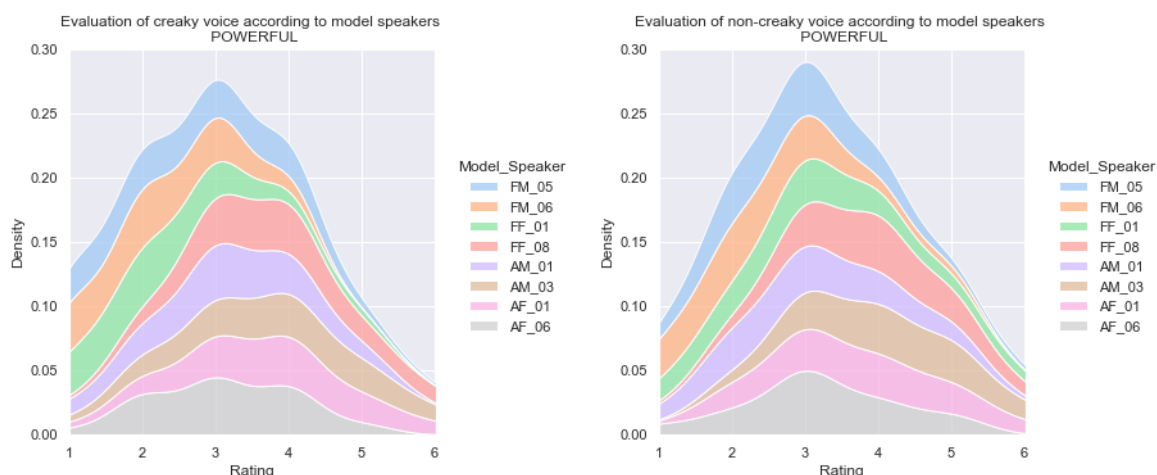


Figure 101: Probability density function of POWERFUL for the different model speakers. Ratings of creaky voice are illustrated on the left panel, those of non-creaky voice on the right panel.

speakers. If we compare ratings of voice qualities within-speaker, the coefficients resulting from our models showed that creaky voice in word-final position, as opposed to non-creaky voice, is associated with:

- small likelihood of being perceived as less powerful for FF_01 ($p < .01$)
- no likelihood of being perceived as less powerful for AF_01, AF_06, AM_03, FF_08, FM_05, FM_06, and FM_05
- no likelihood of being perceived as more powerful for AM_01

The odds of creaky voice in word-final position to be perceived as less powerful are 1.25 times that of non-creaky voice for AF_01 [0.576, 2.705], 1.05 times that of non-creaky voice for AF_06 (95% CI: [0.482, 2.30]), 1.38 times that of non-creaky voice for AM_03 [0.636, 2.993], 3.37 times that of non-creaky voice for FF_01 [1.503, 7.775], 1.19 times that of non-creaky voice for FF_08 [0.545, 2.589], 1.11 times that of non-creaky voice for FM_05 [0.513, 2.417], 1.22 times that of non-creaky voice for FM_06 [0.561, 2.682], holding all variables constant. The odds of creaky voice in word-final position to be perceived as more powerful are 0.56 times that of non-creaky voice for AM_01 [0.251, 1.218], holding all variables constant.

If we compare evaluation of creaky voice in word-final position depending on model speakers, the interpretation of the coefficients resulting from our models showed that, with AF_01 as reference, creaky voice in word-final position is associated with:

- no likelihood of being perceived as more powerful for AM_03, FF_08
- no likelihood of being perceived as less powerful for AF_06
- small likelihood of being perceived as less powerful for AM_01 ($p < .01$), FF_01, FM_05, and FM_06 ($p < .001$)

The odds of creaky voice in word-final position to be perceived as more powerful are 1.21 times when produced by AM_03 [0.541, 2.724], 1.8 times when produced by FF_08 [0.83, 3.922] than when produced by AF_01. The odds of creaky voice in word-final position to be perceived as less powerful are 0.46 times when produced by AF_06 [0.21, 1.009], 0.41 times when produced by AM_01 (95% CI: [0.186, 0.883]), 0.07 times when produced by FF_01 [0.031, 0.16], 0.18 times when produced by FM_05 [0.082, 0.406], 0.11 times when produced by FM_06 [0.047, 0.246] than when produced by AF_01.

3.4.4 Educated

Figure 101 shows whether model speakers were perceived as sounding more educated when producing creaky or non-creaky voice in word-final position. Ratings of creaky voice are displayed on the left panel and those of non-creaky voice on the right panel. Peaks are very similarly distributed across voice qualities, confirming our observations that model speakers sound as educated when producing either creaky and non-creaky voice sound as educated when produced by every model speaker. If we compare ratings of voice qualities within-speaker, the interpretation of the coefficients resulting from our models showed that creaky voice in word-final position, as opposed to non-creaky voice, is associated with:

- no likelihood of being perceived as less educated for FF_01, FF_08, FM_05, and M_06

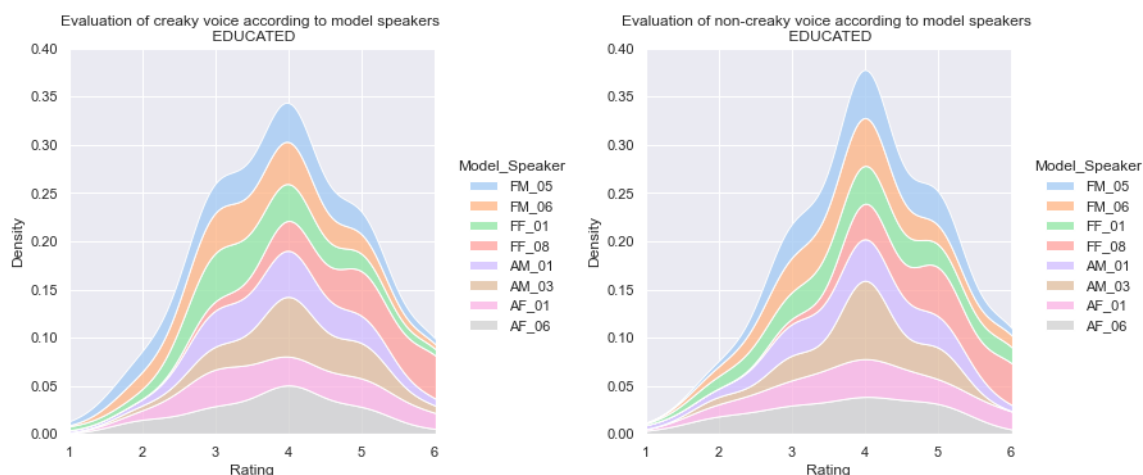


Figure 102: Probability density function of EDUCATED for the different model speakers. Ratings of creaky voice are illustrated on the left panel, those of non-creaky voice on the right panel.

- no likelihood of being perceived as more educated for AF_01, AF_06, AM_01, and AM_03

The odds of speakers using creaky voice in word-final position to be perceived as less educated are 1.42 times that of not using creaky voice for FF_01 [0.642, 3.149], 1.03 times that of not using creaky voice for FF_08 [0.463, 2.277], 1.93 times that of not using creaky voice for FM_05 [0.876, 4.309], 2.18 times that of not using creaky voice for FM_06 [0.994, 4.873], holding all variables constant. The odds of speakers using creaky voice in word-final position to be perceived as more educated are 0.76 times that of not using creaky voice for AF_01 [0.351, 1.654], 0.7 times that of not using creaky voice for AF_06 [0.364, 1.73], 0.7 times that of not using creaky voice for AM_01 [0.318, 1.533], 0.45 times that of not using creaky voice for AM_03 [0.186, 1.056], holding all variables constant.

If we compare evaluation of creaky voice in word-final position depending on model speakers, the interpretation of the coefficients resulting from our models showed that, with AF_01 as reference, creaky voice in word-final position is associated with:

- no likelihood of being perceived as less educated for AF_06, AM_01, AM_03, FM_05, and FM_06

- small likelihood of being perceived as less educated for FF_01 ($p < .01$)
- small likelihood of being perceived as more educated for FF_08 ($p < .001$)

The odds of creaky voice in word-final position to be perceived as more educated are 6.23 times when produced by FF_08 [2.687, 14.561] than when produced by AF_01. The odds of creaky voice in word-final position to be perceived as less educated are 0.71 times when produced by AF_06 [0.311, 1.639], 0.73 times when produced by AM_01 [0.324, 1.633], 0.9 times when produced by AM_03 [0.386, 2.039], 0.29 times when produced by FF_01 [0.126, 0.659], 0.51 times when produced by FM_05 [0.223, 1.171], 0.56 times when produced by FM_06 [0.241, 1.311]) than when produced by AF_01.

In this section we saw that both qualities were overall similarly evaluated, with non-creaky voice slightly more positively than creaky voice. Perceptual judgments differed between- and within-language, with creaky voice being more positively evaluated when produced by American English speakers than by French speakers, and more positively than non-creaky voice in American English but more negatively than non-creaky voice in French. When we included Gender, we observed that creaky voice was more positively evaluated when produced by female than by male American English speakers, although it was more positively rated than non-creaky voice when produced by male American English speakers. There was no real difference in the evaluation of creaky and non-creaky voice when produced by female American English speakers. Creaky voice was also overall more positively evaluated when produced by female French speakers than by male French speakers, although creaky voice was always more negatively judged than non-creaky voice, when either produced by male or female French speakers. Model speaker effect on the evaluation of creaky vs. non-creaky voice revealed a great deal of inter-individual variability in judgment, but more importantly that one particular voice was significantly more positively rated than others, regardless voice quality usage. Conclusion we can draw from this analysis is that listeners might resort to other speech and voice features along with voice quality to positively/negatively judge a voice.

4 Conclusion

In this chapter we provided thorough analyses and results from the three different tasks we conducted: the reading task, the repetition/imitation task, and the rating task. Analysis of creaky voice usage in the reading task revealed more creaky voice occurrences on word-final position when subjects read in English as compared to when they read in French, corroborating the fact that creaky voice usage depends on language status. We did not find any correlation between creaky voice usage in the L1 as compared to that in the L2. Female French learners of English seem to have consciously or unconsciously identified creaky voice as being a feature of English, and have adapted to it due to exposure to the English language (Pillot-Loiseau *et al.*, 2019). Although the number of tokens varies per category in our dataset, we observed significantly more creaky low than high vowels, in both French and English, corroborating previous findings that low vowels are more often creaked than high vowels (Panfili, 2015).

Taken all measurements together, we observed no significant convergence effects from French learners of English towards all model speakers, indicating no real success in the imitation of creaky voice. In the following we only report significant effects. Only vowel duration received overall convergence effects, while H1*, H2*, H1*-A2* and H1*-A3* received overall divergence effects. We observed a language effect on some dimensions, with overall more convergence and less divergence towards/from American English speakers. There was considerable variability as to what dimensions received more convergence or divergence effects across language. Convergence towards American English model speakers was observed on f_0 and PIC, as well as convergence towards French model speakers on CPP. Divergence from both American English and French speakers was observed on H1*, H2*, and H1*-A3*, with greater effect for French model speakers. Subjects also diverged from French model speakers on f_0 and on PIC, and from American English model speakers on CPP. Maintenance was observed towards American English speakers on H1*-H2*, towards French male model speakers on vowel duration, and towards female American English speakers on f_0 and H1*-A2*. Convergence effects were observed towards male French model speakers on f_0 and CPP. Subjects also diverged from that same group on H1*, H2*, H1*-A2*, H1*-A3* and CQ. Convergence was observed towards female French model speakers on duration and CQ. That same group received significant divergence effects on f_0 , H1*, H2*,

H1*-A2*, H1*-A3* and CPP. Subjects converged towards male American English model speakers on duration, f_0 and CQ. They significantly diverged from that group on H1*, H2*, H1-A2*, H1*-A3* and CPP. Significant convergence effects were observed towards female American English speakers on duration, H1*-A3*, and CPP. That same group received divergence on H1*, H2*, and CQ. Cross-gender differences observed within-language showed that subjects converged more towards female than towards male French speakers on vowel duration, and more towards male than towards female American English model speakers on f_0 . H1* and H2* received more divergence from male than from female model speakers, regardless of language. More divergence from male than towards female American English speakers was observed on H1*-A2*. Divergence effects were also observed from male American English model speakers on H1*-A3* and CPP, while convergence was observed for these measure towards their female counterparts. The same pattern was observed for CQ and PIC in French. Including Model speaker as a variable revealed a lot of variation in accommodation patterns towards model speakers. Some model speakers received convergence on dimensions and divergence on others. Overall, each model speaker received divergence on more dimensions than convergence. There was a lot of variability as to what dimensions subjects converged to or diverged from depending on model speakers.

Ratings of creaky vs. non-creaky voice showed that, overall, both voice qualities were similarly evaluated, with non-creaky voice slightly more positively. Creaky voice was only rated as being overall more attractive. Breaking down the data across language, we observed more positive judgment for creaky vs. non-creaky voice when produced by American English speakers, while non-creaky voice was rated more positively than creaky-voice when produced by French speakers. Creaky voice was also more positively rated when produced by American English than by French speakers. The cross-gender analysis we conducted within-language revealed more positive judgment towards female than towards male creaky voice, although it was overall more positively judged than non-creaky voice when produced by male American English (apart from ratings of 'powerful'). There was barely any difference in ratings of creaky vs. non-creaky voice when produced by female American English speakers. Creaky voice only sounded more attractive. When produced by French speakers, creaky voice was more positively rated when produced by female speakers. However, creaky voice was always rated more negatively than non-creaky

voice, whether produced by male or female French speakers. When we compared evaluation of creaky vs. non-creaky voice across model speakers, we observed significantly more positive judgment towards one voice in particular, whether the model would produce creaky or non-creaky voice. The conclusion that we can draw is that listeners' evaluation of a particular voice is probably influenced by other voice and speech parameters than just voice quality. The fact that all traits received more variation in terms of ratings in the creaky voice scenario suggests a great inter-individual variability in how listeners perceive a voice. Although these observations can give us insights as to how creaky voice is perceived across language and gender, our statistical analyses did not result in enough significant results, which is likely to be due to sample size. The likelihood and the odds of one voice quality to be perceived as more/less pleasant, more/less attractive, to sound more/less powerful and more/less educated, were very low, but also differed between- and within-model speakers, depending on features. These observations should, therefore, be carefully interpreted.

CHAPTER VI: General conclusion and discussion

1 Creaky voice

This thesis focuses on the evaluation and accommodation of creaky voice by female French learners of English. Creaky voice, that is sometimes referred to as *creak*, *vocal fry*, *glottalization*, or *laryngealization* in the literature, is said to be produced when the vocal folds vibrate very slowly and spend more time approximated than apart (Gobl & Ní Chasaide Chasaide, 1992, 2010; Johnson, 2011; Podesva & Callier, 2015). Non-modal phonation types like creaky voice can be used in various linguistic ways. They can be varied to produce phonological contrasts, arise as allophonic variants, or be used to mark prosodic boundaries and prominence. Neither American English nor French use creaky voice contrastively. In American English, creaky voice can occur as allophonic variants of modal vowels when they are followed by voiceless stop consonants (Gordon & Ladefoged, 2001). This is due to the fact that voiceless stops are often glottalized in coda position in English. An increase in creak at phrases boundaries has also often been observed, both phrase-initial and phrase-final (Umeda, 1978; Kreiman, 1982; Pierrehumbert & Talkin, 1992; Pierrehumbert, 1995; Dille et al., 1996; Redi & Shattuck-Hufnagel, 2001; Epstein, 2003; Ní Chasaide & Gobl, 2004; Slifka, 2006; Wolk et al., 2012; Podesva, 2013; Garellek, 2014). Pierrehumbert & Talkin (1992) observed higher rates of glottalization on stressed-vowel initial syllables. Garellek (2014) observed an effect of both prominence and phrase-initial position on the presence of word-initial position. He argued that word-initial glottalization may arise from prosodic strengthening, a process by which sounds are more 'strongly' articulated in stronger

positions. Creaky voice does not seem to serve a linguistic function in French apart from being a hesitation marker (Benoist-Lucy & Pillot-Loiseau, 2013).

Variations in voice qualities can also signal paralinguistic information on mood and attitude. In American English, creaky voice was found to be a frequent feature in the speech of female radio newscasters (Dilley *et al.*, 1996). The authors argued that it might have been to take an authoritative stance. This was supported by Lefkowitz & Sicoli (2007) who studied the speech of American college-aged women and found that they were more likely to use creaky voice when having an authoritative attitude. Mendoza-Denton (2007) analysed the speech of Latina gang-affiliated girls in Northern California and found that they would use creaky voice when narrating fight stories. She hypothesized that creaky voice could index toughness.

Voice quality can also serve a sociolinguistic function, to index group membership, to convey different images and personae, or can be used to produce “styling effects” (Pennock, 2015). Esling (1978) found that, in Edinburgh English, the use of creaky voice was associated with a higher social status. Similar findings were observed for Glasgow English (Stuart-Smith, 1999) and for Australian English (Pittam, 1987). In their study of RP and Modified Northern, Henton & Bladon (1988) observed more instances of creaky voice in Modified Northern as compared to RP, and more in the speech of male than in the speech of female speakers, regardless of dialect. They concluded that creaky voice could be regarded as a “robust marker of male speech” and was mostly used to sound “hyper masculine” when employed by male speakers of Modified Northern. Creaky voice has been observed as occurring increasingly in the speech of young female American speakers these past few years (Lefkowitz & Sicoli, 2007; Yuasa, 2010; Podesva, 2013). Conflicting results as to how they were perceived when producing creaky voice were obtained depending on studies. Yuasa (2010) found that they were perceived as being urban well-educated, and to be highly positioned in the social hierarchy. Pennock (2015) studied the speech of three American actresses playing both American and British characters to show that they would deliberately manipulate voice quality to attain specific styling effects. His study revealed that creaky voice was more extensively used in American films and when characters embodied positive stereotypes of femininity. He concluded that desirability depended on cultural setting for creak is considered as desirable in America but less so in Britain. Creaky voice is a frequent component of the voices

of American actresses or known celebrities who act as role models to many young women in America. Gottliebson *et al.* (2007) showed that vocal fry was consistently and deliberately used by many college students. They hypothesized that these college students have either practiced, or perceived this voice quality, and modelled it to match popular figures. Anderson *et al.* (2014) found that vocal fry was negatively perceived by American adults in a labour market context. Both male and female speakers of American English were perceived as being less educated, less trustworthy, less competent, less attractive, and less hireable.

2 The present study

In a previous study we conducted on accommodation of L2 speech, we observed that female French learners of English who had more of an American accent would produce quite a few instances of creaky voice. Many studies have been conducted on linguistic and phonetic accommodation but very few on voice quality adaptation. To the best of our knowledge, this study is the first on accommodation of creaky voice from French learners of English towards both French and American English speakers. Our analysis includes both acoustic and electroglottographic (EGG) measures of convergence and a perceptual evaluation of voice quality. We tested the influence of language (French/American English) and gender (male/female) on creaky voice accommodation and evaluation. We expected more convergence towards American English than towards French creaky voice, and more towards male than towards female model speakers. More positive evaluation towards American English creaky voice was also expected, and more towards female than towards male.

Speech stimuli were elicited from 12 native speakers of both French and American English (3 male French, 3 female French, 3 male American, 3 female American). Audio and electroglottographic signals were recorded simultaneously. We extracted and analysed ten acoustic measures of phonation and two EGG measures. Ten short sentences were produced with either word-final creak or without word-final creak by each model speaker, in their native language. We focused on the last word of the prosodic unit to keep as much balance between the two languages. The last accented word is a monosyllabic word containing either a low or a high vowel. We only used

declarative sentences for they elicit a low tone in both languages.

We first conducted a comparative analysis of creaky voice across language and gender. Many different studies have described the acoustics of American English creaky voice (e.g. Keating & Garellek, 2015; Keating *et al.*, 2015) but little is known as to whether, and how, it differs from French creaky voice. Extracted measurements were compared across all groups. Principal component analyses were performed to observe what variables accounted for most variance across language.

Accommodation was studied from a multi-dimensional perspective. We provided a thorough analysis on multiple acoustic and electroglottographic measures and observed whether female French learners of English would align on the same dimensions to both native speakers of French and American English. Gender influence was analysed within-language. 20 cisgender female native speakers of French aged 20-30, all majoring in English at Université Paris Cité took part in the experiment. The experiment consisted of three different tasks: a reading task, a repetition/imitation task, and a judgment rating task. Subjects first conducted the reading task in which they had to read the sentences corresponding to the stimuli recorded by the model speakers. This served as their baseline production. Subjects conducted the repetition/imitation task after completing the reading task. They were presented with the different auditory stimuli previously recorded. They had been instructed to repeat and imitate as closely as they could the production of the model speakers. Although some speakers appeared confused as to what speech features they were supposed to imitate, no further instructions were given to them. We analysed convergence effects on every acoustic and EGG measurements we made.

Finally, subjects conducted a judgment rating task in which they had to rate their impression of the model speaker's voice they were presented to. They listened to 3 same sentences produced with either word-final creak or without word-final creak produced by 8 model speakers (2 male French, 2 female French, 2 male American, 2 female American). Each voice was rated on a 4 six-point semantic scale including two likeability ratings (pleasant and attractive) and two competence ratings (powerful and educated).

3 Summary of the results and discussion

3.1 Comparative analysis of creaky voice across language and gender

Only CPP and f_0 significantly distinguished between French and American English creaky voice. American English creaky voice was produced with lower f_0 and higher CPP values than French creaky voice. CPP is a measure of aperiodicity. A low CPP value can either indicate creakier or breathier type of phonation (greater ratio of aperiodic to periodic sound in the signal). Although differences were not significant, we also observed higher values of H1*-H2*, H1*-A1*, H1*-A2*, H1*-A3* and PIC, as well as lower CQ value in French creaky voice, indicating that French creaky voice was produced with a less constricted glottis, slower vocal fold closure, and with vocal folds spending less time approximated than apart as compared to American English creaky voice. Several sub-types of creaky voice have been identified in American English (Keating *et al.*, 2015), each having their own set of characteristics. It seems very likely that French creaky voice corresponds to what Keating *et al.* (2015) refer to as non-constricted creak. This would not be surprising for French is a language that favours openness and breathiness, which are both enabled by larynx lowering, rather than constriction (Esling *et al.*, 2019). It might be articulatory less demanding for French people to produce non-constricted creak.

Only f_0 and H2*-H4* significantly distinguished creaky voice across gender, without factoring in language. Men have longer vocal folds than women, resulting in an overall lower fundamental frequency. However, if f_0 only distinguished creaky voice across gender due to physiological differences in men and women, then it would mean that f_0 also distinguished creaky voice across gender and within-language. Our data suggests that it was not the case, indicating that both male and female speakers can reach similar f_0 in the production of creaky voice (Blomgren *et al.*, 1998; Chen *et al.*, 2002). This difference might only result from cross-language differences and inter-individual variability, with both male and female French speakers producing higher f_0 values than their American counterparts.

Cross-gender analyses conducted within-language revealed no significant difference between the production of male and female American English creaky voice. Only PIC significantly distinguished between male and female French creaky voice. Male produced creaky vowels with overall higher PIC values. If PIC does correlate with speed of vocal closure, then it means that vocal

fold closure is greater when creaky voice is produced by male French speakers. Higher PIC values are correlated with a breathier phonation (Esposito, 2012; Keating *et al.*, 2012), suggesting that there was more air leakage in the production of French creaky voice when produced by male speakers. Women have been found to be breathier than men, which is due the fact that male speakers' vocal folds are longer and thicker, which helps closing the glottis completely, shutting of the airflow briefly. The thinner vocal folds of female speakers might never make a complete closure, leading to the presence of a permanent glottal chink, therefore of air leakage (Titze, 1989). Our results suggest that male speakers might have been breathier than female French speakers in our dataset, which is in contradiction with previous studies.

Considerable inter-individual variability was observed in creaky voice production in our dataset. Results of principal component analyses showed that $H1^*-A1^*$, $H1^*-A2^*$, and $H1^*-A3^*$, the three spectral measures that are said to be correlated with vocal fold closure, accounted for most variability in both languages, and particularly in French. There was also significantly more variability in f_0 and $H1^*-H2^*$ in American English creaky voice than in French creaky voice. These observations are in line with the fact that there exists different sub-types of creaky voice in American English which are not implemented acoustically in the same way (e.g. aperiodic creaky voice is not produced with a low f_0 , and non-constricted creaky voice is not produced with low $H1^*-H2^*$). Because these two measures did not account for a lot of variability in French, there might not be as many sub-types of creaky voice existing in French.

Cross-gender analyses conducted within-language showed that there was substantially more variability observed when creaky voice was produced by female speakers than by male speakers of American English. Female American English speakers might produce more sub-types of creaky voice than their male counterparts. There was no striking cross-gender difference in French. These results show that creaky voice usage does not only result from physiological constraints but is a socially-constructed phenomenon. It is more predominant in American English than in French.

Correlation analysis showed that CQ was overall correlated with the same variables in both languages. It was negatively correlated with $H1^*-H2^*$ and f_0 , meaning that, when CQ increases, both of these measures decrease. It is not surprising to observe such correlations. When CQ

increases the vocal folds spend more time approximated than apart, therefore the degree of constriction increases (low $H1^*-H2^*$) and fundamental frequency decreases (slow vibration rate). Negative correlations with $H1^*-A1^*$ and $H1^*-A2^*$ show that the more time the vocal folds spend approximated, the smaller the posterior glottal opening will be (low $H1^*-A1^*$), and the more skewed the shape of the glottal pulse will be (low $H1^*-A2^*$). These results are in line with Childers & Lee's (1991) and Esposito's (2012).

Correlations for PIC went in the opposite direction depending on language. Consistent with Esposito (2012), PIC was negatively correlated with $H1^*-A1^*$, $H1^*-A2^*$ and $H1^*-A3^*$ in French, but positively correlated with those same measures in American English. All these measures are said to be correlates of vocal fold closure in the literature, but correlations going in opposite directions might indicate otherwise. As mentioned by Esposito (2012: 475), who also observed a negative correlation between these measures in her study of White Hmong, "[...] if creaky phonation does indeed have a slower vocal fold closure, then this should also affect the spectral measures $H1^*-A1^*$, $H1^*-A2^*$, $H1^*-A3^*$. These measures work under the principle that faster vocal fold closure excites the higher frequencies of a vowel, making $A1^*/A2^*/A3^*$ greater than $H1^*$. [...] If the vocal folds are vibrating more slowly during creaky phonation and have a slower rate of vocal fold closure, then what is causing the value $A1^*/A2^*/A3^*$ to be greater than that of $H1^*$?" Ge *et al.* (2022) also observed weak correlations between PIC and spectral measures. It does not seem very clear what aspects of speech production the PIC measure really captures and more research is needed to fully understand these findings. PIC and CQ were also positively correlated in French but negatively correlated in American English. Negative correlation between these two measures are consistent with Esposito (2012). She argued that, since creaky voice is produced with vocal folds being close together, they do not need to move as quickly to reach a state of closure. The more contact between the vocal folds, the higher the CQ value, but the lower PIC values. However, the positive correlation between these two measures observed in French creaky voice highlights the fact that more research is needed as to what aspects of phonation PIC is measuring.

This descriptive and comparative analysis of French and American English creaky voice reinforces the idea that phonation types with the same descriptive name differ along several dimen-

sions across language. Our observations are in line with the fact that “language/speaker/recording differences in voice quality are larger than phonation category differences” (Keating *et al.*, 2010: 198). The perspective of producing a combination of different phonation types should, also, further be explored (Esling *et al.*, 2019).

3.2 A comparison across tasks

In this subsection we review our findings relative to our hypotheses presented in Chapter 2.5. We observed a little more convergence towards American English creaky voice and some within-language cross-gender effects on convergence. Creaky voice was overall more positively evaluated when produced by American English speakers than by French speakers. It was more positively evaluated when produced by female speakers within-language. However, creaky voice was always more negatively evaluated than non-creaky voice when produced by French speakers.

3.2.1 Reading task

Results show that female French learners of English produced significantly more word-final creak when reading in English than when reading in French, which corroborates the fact that creaky voice usage depends on language status (Pillot-Loiseau *et al.*, 2019). We did not find any significant correlation between creaky voice usage in the L1 and in the L2 in our data. It seems to be the case that female French learners of English have identified creaky voice as being a feature of English, and have adapted to it due to exposure to the English language (in their life and studies). They might have realised that creaky voice in American English embodied the idea of empowerment, and have adapted to it to convey that same image when speaking English. As we previously saw, phonatory settings in French have been described as nasalized and breathy, but there has been no mention of a possible creaky voice phenomenon (Benoist-Lucy & Pillot-Loiseau, 2013). Creaky voice was, however, also observed in sentence-final position when subjects read in French. Do these manifestations result from the fact that aerodynamic conditions are no longer favourable for modal voice to be used in this position, or do they result from the phenomenon of creaky voice spreading across the ocean? Studies of French creaky voice conducted in more natural settings, that would include male speech, as well as other than just

sentence-final creak occurrences (if any), would provide further results.

3.2.2 Repetition/Imitation task

Inconsistencies were observed across the different measures of convergence we analysed. We observed no overall success in the convergence of creaky voice across the dataset. Only duration displayed significant alignment, while H1*, H2*, H1*-A2*, H1*-A3* received significant divergence effects. Adding Language as an independent variable yielded different results. Subjects converged on more/diverged on less dimensions towards/from American English speakers. Significant divergence effects were observed on H1*, H2*, and H1*-A3*, in both languages, with greater effect for French. Creaky vowels received convergence on f_0 and PIC when produced by American English speakers, but divergence when produced by French speakers. The reverse pattern was observed for CPP. Maintenance was observed on H1*-H2* towards American English speakers while this measure received divergence when repeating French speakers. Cross-gender analyses conducted within-language revealed considerable variability on all dimensions. In American English, both male and female received alignment on duration and divergence on H1*, H2*, H1*-A3*. Male received alignment on f_0 and CQ while female received maintenance on f_0 and divergence on CQ. Female received convergence on CPP and male divergence. Male received divergence on H1*-A2* and female maintenance. In French, female received convergence on duration while male received maintenance. Both male and female received divergence on H1*, H2*, H1*-A2*, and H1*-A3*. Female received divergence on f_0 and CPP while male received convergence. The reverse pattern was observed for CQ. Including Model speaker in our analysis displayed even more variability in accommodation patterns. Subjects would converge/diverge on some dimensions by producing increased/decreased values depending on model speaker. Some models received convergence on some dimensions and divergence on others. There was a lot of variability as to what dimensions received convergence or divergence effects depending on model speaker. Cross-gender analyses revealed a lot of variability as to what dimensions received more convergence/divergence effects. However, no consistent pattern and no significant difference as to whether male or female speakers received aligned on more dimensions were observed within-

language⁶⁷.

Our comparative study of French and American English creaky voice revealed that only CPP and f_0 significantly distinguished between French and American English creaky voice. Creaky voice was produced with higher CPP values and lower f_0 values by American English speakers than by French ones. Accommodation patterns for these measures were reversed across language. Convergence effects were observed on CPP and divergence effects on f_0 towards French models. The reverse pattern was observed when repeating American English speakers. Subjects converged/diverged on these dimensions by producing decreased values. It might be that what subjects perceived more in American English creaky voice was the low fundamental frequency component while it was the breathier kind of quality in French creaky voice. This interpretation would be in line with the fact that subjects diverged on H1*-H2* by producing higher values in French, therefore creaky voice with less glottal constriction. The fact that subjects overall converged on more dimensions towards American English than towards French creaky voice might be due to the fact that there was much more variability in the way creaky voice was produced by American English speakers. More accommodation strategies can arise towards greater acoustic spaces.

This multidimensional analysis of voice quality convergence provided very complex findings that are extremely hard to interpret. Substantial variability as to what dimensions received convergence/divergence depending on groups were observed. One result that needs to be highlighted is the fact that the tendency of subjects to adapt their speech to that of model speakers has proven to be extremely variable. Individual subjects were probably more sensitive to different acoustic variations depending on model speaker, which then influenced convergence effects. In addition, we mentioned that some subjects were confused as to what speech features they were supposed to imitate. They asked me whether they should imitate the model speakers' accent, or the intonation patterns, for instance. No further instructions were provided. Subjects might have directed their attention towards other more salient speech features than voice quality⁶⁸.

⁶⁷Please refer to Tables 40 and 41 in chapter 5.2 for a synthesis of the significant patterns of accommodation observed on acoustic and EGG measures, respectively.

⁶⁸Other speech features might include speech rate (e.g. Pardo *et al.*, 2010), vowel duration (e.g. Zajac, 2013; Lewandowski & Nygaard, 2018), vowel space (e.g. Babel, 2009, 2012; Pardo *et al.*, 2010, 2017).

As mentioned by Kreiman *et al.* (2008: 345), “[...] the listener takes advantage of whatever acoustic information is available to achieve a particular perceptual goal. Which aspects of the signal are important will depend on the task, the characteristics of the stimuli, the listener’s background, perceptual habits, and so on. Given the many kinds of information listeners extract from voice signals, it is not surprising that these characteristics vary from task to task, voice to voice, and listener to listener”. In addition, speakers of languages that contrast phonation types are more accurate at distinguishing different phonation types. Kreiman *et al.* (2010) found that speakers of Mandarin were found to be particularly sensitive to variations in H1-H2. Their perceptual experience with the phonation differences associated with contrastive tone in that language is likely to be responsible for the increased sensitivity observed to changes in harmonic amplitudes. The authors concluded that “there exists a continuum of sensitivity to H1-H2 based on the role this feature plays in the listeners’ native language [and that] native language affects both sensitivity to source characteristics and perceptual strategy” (2010: 592). The sensitivity is going to be lower (yet not inexistent) for speakers of languages that do not contrast phonation types. French speakers’ lack of sensitivity to such changes in the harmonic structure might also account for the minor convergence effects observed in our study.

Adaptation processes might also result more from intrinsic variability than be driven by external social factors. In a recent work, Lee *et al.* (2021) observed that within-speaker variability will actually influence individual speaker adaptability and convergence behaviour. Some types of variance were found to be idiosyncratic, which suggests that voice spaces are not shared across individuals. It might be that greater within-speaker variability resulted in greater convergence effects in our study⁶⁹.

The possible sources of random variability affecting convergence effects appear to be very complex and are certainly numerous. Our results provide further evidence that particular attention should be drawn on individuals in convergence studies, rather than on social groups.

⁶⁹I would always listen briefly to the recordings of the subjects after they conducted the experiment. I realised that those I had perceived as having a more monotonous voice during interactions we had before the experiment were also those who appeared to have a stronger French accent and to sound as poor imitators. Although these observations only result from my own perception, they could corroborate Lee et al.’s (2021) if correlations between these different parameters were actually found.

3.2.3 Judgment rating task

Both non-creaky and creaky voice were similarly evaluated, with non-creaky voice slightly more positively. Including Language as independent variable showed more positive judgment for creaky vs. non-creaky when produced by American English speakers, while non-creaky voice remained more positively rated when produced by French speakers. Creaky voice was also more positively rated when produced by American English speakers than by French speakers. Cross-gender analyses showed that creaky voice was overall more positively rated when produced by female than by male speakers. It was more positively rated when produced by male speakers in American English. There was barely any difference in ratings for creaky vs. non-creaky voice when produced by either male or female French speakers. Creaky voice was always slightly more negatively rated. However, it was more positively rated when produced by female than by male French speakers. Voice quality attitudes can therefore be language-specific and our study shows that creaky voice has more prestige in American English. We previously saw that young adult American women speaking with a creaky voice were perceived as being urban, well-educated (Yuasa, 2010), desirable (Pennock, 2015), or more socially dominant (Borkowska & Pawlowski 2011). Our data shows that creaky voice was also overall rated more positively when produced by female speakers. In addition, more occurrences of creaky voice were observed when subjects were reading in English, and no correlation between creaky voice usage in the L1 and in the L2 was found, indicating that creaky voice usage did not result from idiosyncrasies, but from adaptation towards the L2. We can hypothesize that female French learners of English have adapted to L2 creaky voice to convey that same positive image it conveys in American English, but also maybe to convey the image of empowered language learners. Speaking two or more languages has always been positively perceived and one goal language learners have is to acquire a good pronunciation to eventually sound near-native. Acquiring L2 speech features is seen as an achievement and increases perceived proficiency. Kallio *et al.* (2022) investigated the role of creaky voice to predict fluency and proficiency of spontaneous L2 Finnish speech. Creaky voice serves as a cue for turn-taking and phrase-boundaries in Finnish (Lehiste, 1965). Creaky voice was found to contribute to both fluency and oral proficiency. It might be that French learners of English also adapted to L2 creaky voice in order to sound more native-like.

We saw that not all types of creaky voice are produced with a low f_0 and that this measure significantly distinguished French and American English creaky voice in our dataset. American English creaky voice was produced with significantly lower f_0 values. Hodges-Simeon *et al.* (2010) found that men speaking with a low f_0 positively influenced the perception of attractiveness by women. Creaky voice when produced by male American English speakers had the lowest f_0 values in our dataset. It was also perceived as being more attractive than non-creaky voice when produced by this group, which corroborates Hodges-Simeon *et al.*'s (2010) observations. Our results are in line with the fact that American men with lower and deeper voices are perceived more positively compared to those with higher-pitched voices (Puts, 2005; Puts *et al.*, 2007). However, Biemans (2000) observed that breathy voice was associated with “sexy voice” and Esling *et al.* (2019) argued that female speakers might also slightly prefer breathy male voices, in order to counteract the impression of a large body size that could indicate aggressiveness. Breathiness has been observed as being part of the French phonatory settings (Benoist-Lucy & Pillot-Loiseau, 2013) and non-creaky voice was more positively rated than creaky voice when produced by male French speakers. We suggest that American English male speakers are more positively perceived when producing creaky voice and French male speakers are more positively perceived when producing breathy voice. This was confirmed by two participants who, after completing the judgment rating task, told me that this task made them realise that they preferred *les voix calmes et suaves*, calm and sweet voices, which seemed to be opposite as to what creaky voice sounds like: “running a stick along a fence, or slowly opening a door with creaky hinges” (Biemans, 2000: 27).

Analysing evaluation of creaky vs. non-creaky voice across model speakers show that, depending on traits, creaky voice was either more positively or more negatively rated. These observations corroborate those made by Callier (2013) who also found inconsistencies on the effects which creaky voice had on evaluations on speakers, meaning that the indexical content of creaky voice is not static. We also observed significantly more positive ratings for one voice in particular, regardless of voice quality, which implies that speakers relate to other perceptual features than just voice quality when judging a voice. Perception and judgments are based on a combination of speech features that interact with one another in unique ways (Parker & Borrie,

2017). Papcun *et al.* (1989) argued that listeners who have had a life-long experience with voice develop central category constituents for vocal quality that they will use to judge or remember voices they hear. These categories deriving from perceptual experience, listeners' background will affect their perceptual strategy when judging voices.

4 Methodological issues

Nothing is perfect and this thesis has several flaws. One thing I have learned is that one will inevitably face both expected and unexpected difficulties when conducting experimental studies, especially in studies involving human speech. No speaker says the same thing in exactly the same way twice, which is part of the beauty of language, but which can become problematic in controlled studies. Some voices can also be very unsteady in quality and it has to be borne in mind that individual speakers might produce creaky voice in different ways, which introduces confounding factors. We reported studies in which the rate and proportion of creaky voice were found to vary with several factors such as dialects, gender, location in the utterance, prosodic structures, etc. For these reasons we used a standardized procedure in which we tried to maintain a tight control between all of our variables. We wanted each model speaker to produce the same sentences with either word-final creak or without creak, hence to get the same number of low and high vowels produced in both qualities by each model talker. We quickly realised that some speakers would not produce enough occurrences of creaky voice, while others would produce too many. We eventually had to accept the fact that we would have to use different sentences across speakers, as well as reducing the number of model speakers and sentences, if we wanted to move forward. Extreme and unavoidable intra- and inter-individual variability resulted in an unequal representation of vowel categories in our stimuli set: creating a perfectly balanced design was near impossible. Additional but related difficulties are added in cross-linguistic studies. Segments of interest have to respect a certain number of linguistic constraints (e.g. phonological environment, position in the sentence, etc.) to enable comparison across language. This considerably reduced the possibilities to come up with new words. It is also important to keep in mind that there may be considerable dynamic variations within spoken utterances in any voice quality⁷⁰. Recruiting

⁷⁰It also does not necessarily mean that creaky voice did not occur in other positions in the recorded sentences.

subjects who would fit all criteria appeared more challenging than we thought it was. We were unable to test all competences that would assess for general proficiency level, and more specifically we were unable to test speaking proficiency. Although subjects had a consistent level on the competences tested, their speaking abilities might have differed. All subjects read and repeated the same sentences as those produced by model talkers. Unfortunately, some subjects failed to repeat some sentences, resulting in an uneven distribution of tokens per subject. Of course the pandemic we went through did not help. The first lockdown was announced on the very first day I started recording the French model speakers. Experiments were then suspended for 6 months and then a very strict protocol had to be respected. All these difficulties we encountered restricted us from analysing a bigger sample. The use of different methods, speakers, or measures across studies, may yield different interpretations of the same patterns.

5 Future research

In this section we provide ideas and advice for future research studies on voice quality accommodation. Further analyses conducted on our data could provide answers to questions from section 5.1, and partial answers to questions from section 5.2 (additional data should be collected to yield better results). Depending on the perspective of research, other corpora should be built to answer all other questions from section 5.3.

5.1 With current data

1. We extracted and analysed all occurrences of word-final creak we observed in this study. Results seem to show that there exist more types of sub-types of creaky voice in American English than in French. Identifying and coding all sub-types of creaky voice (Keating *et al.*, 2015) among the creaky voice occurrences we extracted might yield better results. It is possible that individual acoustic characteristics subsumed in sub-types of creaky voice affect how listeners perceive and accommodate to these different sub-types.
2. We conducted a language proficiency test to select participants that all had an advanced level. Subjects self-reported their speaking proficiency and degree of accentedness (Dmitrieva

- et al.*, 2020) because their speech could not be tested. There is a clear separation between pronunciation and other L2 skills. Neufeld (1987), for example, found no correlation between the ratings of pronunciation skills and general language aptitude tests. More control in speaking proficiency and providing measurements of phonetic skills would improve our understanding of the relationship between 'phonetic talent' (Lewandowski & Jilka, 2019) and phonetic convergence, and its effect on the acquisition of the L2 sound system.
3. In the language background questionnaire they had to fill out, subjects were asked whether they thought they had more of an American accent or more of a British accent, and whether they preferred American accents or British accents. Testing the influence of preferred accent on creaky voice accommodation could yield interesting results.
 4. Lee *et al.* (2021) recently observed that within-speaker variability will influence individual speaker adaptability and convergence behaviour. Analysing individuals sharing similar phonetic spaces rather than analysing the effect of social groups would provide more controlled data in convergence studies.
 5. Following Kreiman *et al.* (2020), further studies on voice quality should include both formant frequencies and bandwidths as part of the psychoacoustic model of voice quality. This model argues that vocal tract filter has effects on the perception of voice quality. The source and vocal tract functions cannot be separated: "Speakers must adjust source and filter jointly if they are to simultaneously achieve both voice quality and vowel quality goals. This is inconsistent with the distinction between narrow and broad definitions of voice, and suggests that very narrow definitions of voice quality may be untenable" (Kreiman *et al.*, 2020: 464).
 6. Including non-creaky vowels in the convergence analysis might provide interesting results regarding a possible effect of voice quality on accommodation.
 7. Chanclu *et al.* (2021) conducted an automatic classification of phonation types in spontaneous speech. Two binary neural network-based classifiers were used in their study: one to detect modal and non-modal vowels, the other to classify non-modal vowels into creaky and breathy types. Using a classifier-based approach with our data would provide additional

knowledge about the acoustics of American English vs. French creaky voice, or creaky voice in English produced by either native or non-native speakers.

5.2 With additional data

1. Non-modal phonation does not necessarily occur over the entire course of the vowels. Esposito (2012), for instance, found that one of the measures tested distinguished the three phonation types she analysed at the same time point, emphasizing the importance of measuring phonation at various points within a vowel. We made measurements over the entire vowel as well as averaged over each third of the vowel to observe any difference in where the creaky portion was located in American English and French creaky vowels. We were not able to include measurements taken at point 1 and 3 for some CQ and PIC values were output as 0. Primary observations seem to show that creaky phonation occurs particularly at the end of vowels. Further research should include more data to avoid such issues.
2. It has been observed that liking a person generally results in more imitation of that person's behaviour (Stel *et al.*, 2008), but also that imitating a person during interactions increases liking as well as affiliation between conversation partners (Chartrand & Bargh, 1999; Stel *et al.*, 2010). Borrie & Delfino (2017) observed that young adult female American speakers employed significantly more instances of creaky voice when conversing with a partner who exhibited significant creaky voice, as compared to when conversing with a partner exhibiting quantifiably less instances of creaky voice. Greater similarity between interaction partners in their use of creaky voice resulted in a higher score of communicative efficiency and enjoyment. Correlations between evaluations of voice quality (positive/negative) and convergence effects should be further explored.
3. Davidson (2019) analysed the effect of gender, pitch, and utterance type on the identification of creaky voice. Stimuli were either full sentences or sentence-final fragments, produced in either modal voice, with partial creak, or fully creaky. Female American English speakers were identified as being creakier than male speakers, although the effect was not as strong as that of the environment in which creak was produced. More precisely, listeners were more likely to identify creak when the whole utterance was creaky. She argued

(2019: 20) that “overt identification of creaky voice may be more likely to occur when it is unexpectedly found throughout the entire utterance, or is being used other than in its prosodic function of signalling the end of the utterance”. Creak is harder to detect when it is expected in this position, and listeners might treat the different functions of creaky voice in various ways. We only analysed word-final creak to retain as much balance as we could between the two languages. Analysing instances of creaky voice occurring in other positions might yield completely different results.

5.3 With new data

1. Lambert *et al.* (1966) would use stimuli produced by the same speaker in different languages in which they were equally proficient. This technique was developed to uncover language users’ evaluations of language varieties. Bilingual speakers could be used in studies on cross-linguistic perception of non-contrastive phonation to really expose the effect of language attitude on non-modal phonation.
2. Using two different scales, one likeability rating scale (e.g. friendly, likeable, social, etc.), and one competence rating scale (e.g. successful, self-confident, intelligent, etc.) would provide additional information regarding the indexical content of creaky voice (or any other voice quality) across language and gender.
3. More convergence on low vowels, and especially in the F1 dimension has been observed in the literature (e.g. Babel, 2012; Burin & Ballier, 2017). Low vowels also seem to be more subject to creak than high vowels (Panfili, 2015). Unfortunately, we did not succeed in collecting an even number of tokens per vowel category to test the effect of vowel category on creaky voice convergence. Future research should include balanced sample between vowels categories to better interpret the relationship between vowel quality and voice quality. As mentioned by Epstein (2002: 41): “It is possible that each vowel possesses its own intrinsic voice quality. [M]ovement of the jaw can affect the position of the larynx and tension in the vocal folds. In other words, vowel height and frontness could potentially affect the glottal setting of the vowel”. This echoes question 5 above.

4. Some speakers confessed that it was difficult not to be influenced by the semantic content of sentences in the judgment rating task. Neutral semantic content sentences should be favoured in voice evaluation tasks. One could also use single words excised from the stimulus sentence and pasted into a neutral carrier phrase.

Bibliography

- Abberton, E., A. Parker, and A. Fourcin (1977). Speech improvement in deaf adults using laryngograph displays. In *Papers from the research conference on speech analyzing aids for the deaf*, pp. 172–188. Washington, DC.
- Abdelli-Beruh, N. B., L. Wolk, and D. Slavin (2014). Prevalence of Vocal Fry in Young Adult Male American English Speakers. *Journal of Voice* 28(2), 185–190.
- Abercrombie, D. (1968). Paralanguage. *International Journal of Language & Communication Disorders* 3(1), 55–59.
- Abitbol, J., P. Abitbol, and B. Abitbol (1999). Sex hormones and the female voice. *Journal of Voice* 13(3), 424–446.
- Anaconda Inc. (2020). *Anaconda Software Distribution*. Also available as: www.docs.anaconda.com.
- Anderson, A. H., M. Bader, E. G. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. S. Thompson, and R. Weinert (1991). The Hrc Map Task Corpus. *Language and Speech* 34(4), 351–366.
- Anderson, R. C., C. A. Klostad, W. J. Mayew, and M. Venkatachalam (2014). Vocal Fry May Undermine the Success of Young Women in the Labor Market. *PLoS ONE* 9(5), 1–8.
- Andreeva, B., G. Demenko, J. Bachan, I. Gessinger, K. Jankowska, and B. Möbius (2021). Phonetic convergence evaluation based on fundamental frequency variability. In *Proceedings of ESSV*, pp. 1–8. Berlin, Germany.

- Aronovitch, C. D. (1976). The Voice of Personality: Stereotyped Judgments and their Relation to Voice Quality and Sex of Speaker. *The Journal of Social Psychology* 99(2), 207–220.
- Askenfelt, A., J. Gauffin, J. Sundberg, and P. Kitzing (1980). A Comparison of Contact Microphone and Electroglottograph for the Measurement of Vocal Fundamental Frequency. *Journal of Speech, Language, and Hearing Research* (23), 258–273.
- Austin, W. M. (1965). Some social aspects of paralanguage. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 11(1), 31–39.
- Avelino, H. (2010). Acoustic and Electroglottographic Analyses of Nonpathological, Nonmodal Phonation. *Journal of Voice* 24(3), 270–280.
- Babel, M. (2009). Selective Vowel Imitation in Spontaneous Phonetic Accommodation. *UC Berkeley Phonology Lab Annual Report*, 163–194.
- Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society* (39), 437–456.
- Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* (40), 177–189.
- Babel, M., G. McGuire, S. Walters, and A. Nicholls (2014). Novelty and social preference in phonetic accommodation. *Laboratory Phonology* 5(1), 123–150.
- Baken, R. J. (1992). Electroglottography. *Journal of Voice* 6(2), 98–110.
- Baken, R. J. and R. F. Orlikoff (2000). Lanrygeal Function. In *Clinical Measurement of Speech and Voice (Second Edition)*, pp. 393–451. USA: Singular Thomson Learning.
- Bayard, D., C. Gallois, A. Weatherall, and J. Pittam (2001). Pax Americana? Accent attitudinal evaluations in New Zealand, Australia and America. *Journal of Sociolinguistics* 5(1), 22–49.
- Bell, A. (2001). Back in style: Reworking audience design. In P. Eckert and J. R. Rickford (Eds.), *Style and Sociolinguistic Variation*, pp. 139–169. Cambridge: Cambridge University Press.

-
- Belotel-Grenié, A. and M. Grenié (2004). The creaky phonation and the organization of Chinese discourse. In *Proceedings of the International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*, pp. 5–8. Beijing, China.
- Benoist-Lucy, A. and C. Pillot-Loiseau (2013). The influence of language and speech task upon creaky voice use among six young American women learning French. In *Proceedings of Interspeech*, pp. 2395–2399. Lyon, France.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, pp. 13–34. Timonium, MD: York Press.
- Best, C. T. and M. D. Tyler (2007). Nonnative and second-language speech perception. In M. J. Munro and O.-S. Bohn (Eds.), *Language Experience in Second Language Speech Learning. In honor of James Emil Flege*, pp. 13–34. Amsterdam: John Benjamins.
- Bickley, C. (1982). Acoustic analysis and perception of breathy vowels. *Speech communication group working papers*, 71–82.
- Biemans, M. (2000). *Gender variation in voice quality*. Ph. D. thesis, Radboud University, Nijmegen, The Netherlands.
- Bilous, F. R. and R. M. Krauss (1988). Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads. *Language & Communication* 8(3/4), 183–194.
- Birdsong, D. (2006). Age and Second Language Acquisition and Processing: A Selective Overview. *Language Learning* 56(1), 9–49.
- Blankenship, B. (2002). The timing of nonmodal phonation in vowels. *Journal of Phonetics* (30), 163–191.
- Blomgren, M., Y. Chen, M. L. Ng, and H. R. Gilbert (1998). Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers. *The Journal of the Acoustical Society of America* 103(5), 2649–2658.

- Boersma M. and Weenink D. (2020). *Praat: doing phonetics by computer (6.1.51)*. Also available as: www.praat.org.
- Borkowska, B. and B. Pawlowski (2011). Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour* 82(1), 55–59.
- Borrie, S. A. and C. R. Delfino (2017). Conversational Entrainment of Vocal Fry in Young Adult Female American English Speakers. *Journal of Voice* 31(4), 25–32.
- Bourhis, R. Y. and H. Giles (1977). The language of intergroup distinctiveness. In *Language, ethnicity, and intergroup relations*, pp. 119–135. London: Academic Press.
- Branigan, H. P., M. J. Pickering, and Cleland (2000). Syntactic co-ordination in dialogue. *Cognition* 75(2), B13–B25.
- Brown, B. L., W. J. Strong, and A. C. Rencher (1974). Fifty-four voices from two: The effects of simultaneous manipulations of rate, mean fundamental frequency, and variance of fundamental frequency on ratings of personality from speech. *The Journal of the Acoustical Society of America* (55), 313–318.
- Brown, P. and S. Levinson (1987). *Politeness: Some Universals in Language Use*. Cambridge: Cambridge University Press.
- Bulatov, D. (2009). The Effect of Fundamental Frequency on Phonetic Convergence. *UC Berkeley Phonology Lab Annual Report*, 404–434.
- Burin, L. (2018). Accommodation of L2 Speech in a Repetition Task: Exploring Paralinguistic Imitation. *Research in Language* 16(4), 377–406.
- Burin, L. and N. Ballier (2017). Accommodation in learner corpora: A case study in phonetic convergence. *Anglophonia* (24), 1–18.
- Burkhardt, F. and W. F. Sendlmeier (2000). Verification of acoustical correlates of emotional speech using formant-synthesis. In R. Cowie, E. Douglas-Cowie, and M. Schröder (Eds.), *Proceedings of the ISCA Workshop on Speech and Emotion: A Conceptual Framework for Research*, pp. 151–156. Queen’s University, Belfast.

-
- Byrne, D. (1971). *The Attraction Paradigm*. New-York: Academic Press.
- Byrne, D. (1997). An Overview (And Underview) of Research and Theory Within the Attraction Paradigm. *Journal of Social and Personal Relationships* 14(3), 417–431.
- Callier, P. (2010). Voice quality, rhythm and valorized feminities. In *Sociolinguistics Symposium* 18, Southampton, UK.
- Callier, P. R. (2013). *Linguistic Context and the Social Meaning of Voice Quality Variation*. Ph. D. thesis, Georgetown University, Washington, DC.
- Carlson, R., K. Gustafson, and E. Strangert (2006). Cues for hesitation in speech synthesis. In *Proceedings of Interspeech*, pp. 1300–1303. Pittsburgh, USA.
- Catford, J. (1964). Phonation types: The classification of some laryngeal components of speech production. In D. Abercrombie, D. B. Fry, P. MacCarthy, N. C. Scott, and J. L. M. Trim (Eds.), *In Honour of Daniel Jones*, pp. 26–37. London: Longmans.
- Chanclu, A., I. Ben Amor, C. Gendrot, E. Ferragne, and J.-F. Bonastre (2021). Automatic Classification of Phonation Types in Spontaneous Speech: Towards a New Workflow for the Characterization of Speakers' Voice Quality. In *Interspeech 2021*, pp. 1015–1018. Brno, Czech Republic.
- Chartrand, T. L. and J. A. Bargh (1999). The Chameleon Effect: The Perception-Behavior Link and Social Interaction. *Journal of Personality and Social Psychology* 76(6), 893–910.
- Chen, Y., M. P. Robb, and H. R. Gilbert (2002). Electroglottographic Evaluation of Gender and Vowel Effects During Modal and Vocal Fry Phonation. *Journal of Speech, Language, and Hearing Research* 45(5), 821–829.
- Chevalier, F. (2021). *De l'accommodation linguistique au changement phonétique: variabilité idiolectale et variation dialectale dans l'anglais parlé à Glasgow (XX^e)*. Ph. D. thesis, Université de Poitiers, France.
- Chevrie-Muller, C. (1967). Contribution à l'étude des traces glottographiques chez l'adulte normal. *Revue de Laryngologie* (88), 227–244.

- Childers, D., A. Smith, and G. Moore (1984). Relationships between Electroglottograph, Speech, and Vocal Cord Contact. *Folia Phoniatrica et Logopaedica* 36(3), 105–118.
- Childers, D. G. (2000). *Speech processing and synthesis toolboxes*. New-York: John Wiley & Sons.
- Childers, D. G. and C. K. Lee (1991). Voice quality factors: Analysis, synthesis, and perception. *The Journal of the Acoustical Society of America* 90(5), 2394–2410.
- Coles-Harris, E. H. (2017). Perspectives on the motivations for phonetic convergence. *Language and Linguistics Compass* 11(12), 1–33.
- Collins, B., I. M. Mees, and P. Carley (2019). *Practical English Phonetics and Phonology. A Resource Book For Students (Fourth Edition)*. New-York: Routledge.
- Colton, R. H. and E. G. Conture (1990). Problems and pitfalls of electroglottography. *Journal of Voice* 4(1), 10–24.
- Crystal, D. (1997). *A Dictionary of Linguistics and Phonetics (4th Edition)*. Oxford: Blackwell Publishers.
- Danes, P. B. and E. N. Pinson (1993). *The Speech Chain: The Physics and Biology of Spoken Language (Second Edition)*. Oxford: W.H. Freeman and Company.
- Davidson, L. (2019). The Effects of Pitch, Gender, and Prosodic Context on the Identification of Creaky Voice. *Phonetica*, 1–28.
- Davison, D. (1991). An acoustic study of so-called creaky voice in Tianjin Mandarin. *UCLA Working Papers in Phonetics* (78), 50–57.
- De Jong, N. and T. Wempe (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods* 41(2), 385–390.
- de Krom, G. (1993). A Cepstrum-Based Technique for Determining a Harmonics-to-Noise Ratio in Speech Signals. *Journal of Speech and Hearing Disorder* (36), 254–266.

-
- Di Cristo, A. (1999). Vers une modélisation de l'accentuation du français: première partie. *French Language Studies* (9), 143–179.
- Dias, J. W. and L. D. Rosenblum (2016). Visibility of speech articulation enhances auditory phonetic convergence. *Attention, Perception, & Psychophysics* 78(1), 317–333.
- Dilley, L., S. Shattuck-Hufnagel, and M. Ostendorf (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24(4), 423–444.
- Dmitrieva, O., A. Jongman, and J. A. Sereno (2020). The Effect of Instructed Second Language Learning on the Acoustic Properties of First Language Speech. *Languages* 5(44), 1–33.
- Drugman, T., J. Kane, and C. Gobl (2013). Data-driven detection and analysis of the patterns of creaky voice. *Computer Speech and Language* 28(5), 1–39.
- Dufour, S. and N. Nguyen (2013). How much imitation is there in a shadowing task? *Frontiers in Psychology* 4(346), 1–7.
- Edmondson, J. A. and J. H. Esling (2006). The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies. *Phonology* 23(2), 157–191.
- Eisikovits, E. (1987). Sex differences in inter-group and intra-group interaction among adolescents. In A. Pauwels (Ed.), *Women and Language in Australiand and New Zealand Society*, pp. 45–58. Sydney: Australian Professional Publications.
- Epstein, M. (2002). *Voice Quality and Prosody in English*. Ph. D. thesis, University of California, Los Angeles.
- Epstein, M. (2003). Quality and Prosody in English. In *Proceedings of the 15th International Congress of Phonetic Sciences*, pp. 2405–2408. Barcelona, Spain.
- Esling, J. (1978). The identification of features of voice quality in social groups. *Journal of the International Phonetic Association* 8(1-2), 18–23.
- Esling, J. H., S. R. Moisik, A. Benner, and L. Crevier-Buchman (2019). *Voice Quality: The Laryngeal Articulator Model*. Cambridge Studies in Linguistics 162. Cambridge University Press.

- Esling, J. H. and R. F. Wong (1983). Voice Quality Settings and the Teaching of Pronunciation. *TESOL Quaterly* 17(1), 89–95.
- Esposito, C. M. (2004). Santa Ana del Valle Zapotec phonation. *UCLA Working Papers in Phonetics* (103), 71–105.
- Esposito, C. M. (2006). *The effects of linguistic experience on the perception of phonation*. Ph. D. thesis, UCLA, Los Angeles.
- Esposito, C. M. (2012). An acoustic and electroglottographic study of White Hmong tone and phonation. *Journal of Phonetics* 40(3), 466–476.
- Fant, G. (1970). *Acoustic Theory of Speech Production (Second Edition)*. La Haye, Pays-Bas: Mouton.
- Fischer-Jorgensen, E. (1967). Phonetic Analysis of Breathy (Murmured) Vowels in Gujarati. *Indian Linguistics* (28), 71–139.
- Flege, J. and O.-S. Bohn (2021). The revised speech learning model (slm-r). In R. Wayland (Ed.), *Second Language Speech Learning. Theoretical and Empirical Progress*, pp. 3–83. Cambridge University Press.
- Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics* (15), 47–65.
- 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*, pp. 233–277. Timonium, MD: York Press.
- Flege, J. E., M. J. Munro, and I. R. A. MacKay (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America* 97(5), 3125–3134.
- Fourcin, A. and E. Abberton (1976). The laryngograph and the Voiscope in speech therapy. In E. Loebell (Ed.), *Proceedings of the 16th International Congress of Logopedics and Phoniatics*, Basel, Switzerland, pp. 116–122.

-
- Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *The Journal of the Acoustical Society of America* 99(3), 1730–1741.
- Gallois, C. and V. J. Callan (1988). Communication accommodation and the prototypical speaker: Predicting evaluations of status and solidarity. *Language & Communication* 8(3-4), 271–283.
- Garellek, M. (2014). Voice quality strengthening and glottalization. *Journal of Phonetics* (45), 106–113.
- Garellek, M. (2019). The phonetics of voice. In W. Katz and P. Assmann (Eds.), *The Routledge Handbook of Phonetics*, pp. 75–107. London: Routledge.
- Garellek, M. and P. Keating (2011). The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association* 41(2), 185–205.
- Garellek, M. and S. Seyfarth (2016). Acoustic differences between English /t/ glottalization and phrasal creak. In *Proceedings of Interspeech*, pp. 1054–1058. San Francisco, USA.
- Ge, C., W. Xu, W. Gu, and P. Mok (2022). An electroglottographic study of phonation types of tones in Suzhou Wu Chinese. In *Proceedings of Speech Prosody*, pp. 570–574. Lisbon, Portugal.
- Gerratt, B. R. and J. Kreiman (2001). Toward a taxonomy of nonmodal phonation. *Journal of Phonetics* (29), 365–381.
- Gick, B., I. Wilson, and D. Derrick (2013). *Articulatory Phonetics*. Malden, MA: Wiley-Blackwell.
- Giles, H. (1973). Accent mobility: A model and some data. *Anthropological Linguistics* (15), 87–105.
- Giles, H., N. Coupland, and J. Coupland (1991). Accommodation Theory: Communication, Context and Consequence. In J. Coupland, N. Coupland, and H. Giles (Eds.), *Contexts of Accommodation: Developments in Applied Sociolinguistics*, pp. 1–68. Cambridge: Cambridge University Press.

- Giles, H. and P. Johnson (1987). Ethnolinguistic identity theory: a social psychological approach to language maintenance. *International Journal of the Sociology of Language* 1987(68), 69–99.
- Gobl, C. (1988). Voice source dynamics in connected speech. *STL-QPSR* 29(1), 123–159.
- Gobl, C. (1989). A preliminary study of acoustic voice quality correlates. *STL-QPSR* 30(4), 9–22.
- Gobl, C. and A. Ní Chasaide (1988). The effects of adjacent voiced/voiceless consonants on the vowel voice source: a cross language study. *STL-QPSR* 29(2/3), 23–59.
- Gobl, C. and A. Ní Chasaide (1992). Acoustic characteristics of voice quality. *Speech Communication* (11), 481–490.
- Gobl, C. and A. Ní Chasaide (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication* 40(1-2), 189–212.
- Gobl, C. and A. Ní Chasaide (2010). Voice Source Variation and Its Communicative Functions. In J. L. William, J. Hardcastle and F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences: Second Edition*, pp. 378–423. Wiley.
- Goldinger, S. D. (1998). Echoes of Echoes? An Episodic Theory of Lexical Access. *Psychological Review* 105(2), 251–279.
- Goldinger, S. D. and T. Azuma (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review* 11(4), 716–722.
- Gordon, M. and A. Applebaum (2010). Acoustic correlates of stress in Turkish Kabardian. *Journal of the International Phonetic Association* 40(1), 35–58.
- Gordon, M. and P. Ladefoged (2001). Phonation types: a cross-linguistic overview. *Journal of Phonetics* (29), 383–406.
- Gottliebson, R. O., L. Lee, B. Weinrich, and J. Sanders (2007). Voice Problems of Future Speech-Language Pathologists. *Journal of Voice* 21(6), 699–704.

-
- Greenwald, A. G., D. E. McGhee, and J. L. K. Schwartz (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology* (74), 1464–1480.
- Gričić, T. and C. Nilep (2004). When Phonation Matters: The Use and Function of *yeah* and Creaky Voice. *Colorado Research in Linguistics* (17), 1–11.
- Hall, K. (1995). Lip Service on the Fantasy Lines. In M. Bucholtz and K. Hall (Eds.), *Gender Articulated: Language and the Socially Constructed Self*, pp. 183–217. New-York: Routledge.
- Hanson, H. M. (1997). Glottal characteristics of female speakers: Acoustic correlates. *The Journal of the Acoustical Society of America* 101(1), 466–481.
- Hanson, H. M. and E. S. Chuang (1999). Glottal characteristics of male speakers: Acoustic correlates and comparison with female data. *The Journal of the Acoustical Society of America* 106(2), 1064–1077.
- Hanson, H. M., K. N. Stevens, H.-K. J. Kuo, M. Y. Chen, and J. Slifka (2001). Towards models of phonation. *Journal of Phonetics* 29(4), 451–480.
- Henrich, N. (2001). *Etude de la source glottique en voix parlée et chantée : modélisation et estimation, mesures acoustiques et électroglottographiques, perception*. Ph. D. thesis, Université Pierre et Marie Curie (Paris VI), Paris.
- Henrich, N., C. d’Alessandro, B. Doval, and M. Castellengo (2004). On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *The Journal of the Acoustical Society of America* 115(3), 1321–1332.
- Henrich, N., C. d’Alessandro, B. Doval, and M. Castellengo (2005). Glottal open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency. *The Journal of the Acoustical Society of America* 117(3), 1417–1430.
- Henton, C. G. and R. Bladon (1988). Creak as a sociophonetic marker. In L. M. H. . C. N. Li (Ed.), *Language, speech and mind: studies in honour of Victoria A. Fromkin*, pp. 3–29. London: Routledge.

- Hillenbrand, J. H., R. A. Cleveland, and R. L. Erickson (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research* (37), 769–778.
- Hirose, H. (1997). Investigating the physiology of laryngeal structures. In W. J. Hardcastle and J. Laver (Eds.), *The Handbook of Phonetic Sciences*, pp. 116–136. Oxford: Blackwell.
- Hirst, D. and A. Di Cristo (1998). A survey of intonation systems. In *Intonation Systems. A Survey of Twenty Languages*, pp. 1–45. Cambridge: Cambridge University Press.
- Hodges-Simeon, C. R., S. J. C. Gaulin, and D. A. Puts (2010). Different Vocal Parameters Predict Perceptions of Dominance and Attractiveness. *Human Nature* (21), 406–427.
- Hollien, H., P. Moore, R. W. Wendahl, and J. F. Michel (1966). On the Nature of Vocal Fry. *Journal of Speech and Hearing Research* 9(2), 245–247.
- Holmberg, E. B., R. E. Hillman, J. S. Perkell, P. C. Guiod, and S. L. Goldman (1995). Comparisons Among Aerodynamic, Electrolaryngographic, and Acoustic Spectral Measures of Female Voice. *Journal of Speech, Language, and Hearing Research* 38(6), 1212–1223.
- Howard, D. M. (1995). Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers. *Journal of Voice* 9(2), 163–172.
- Iseli, M., Y.-L. Shue, and A. Alwan (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *The Journal of the Acoustical Society of America* 121(4), 2283–2295.
- Ishi, C. T., H. Ishiguro, and N. Hagita (2005). Proposal of acoustic measures for automatic detection of vocal fry. In *Proceedings of Interspeech*, pp. 481–484. Lisbon, Portugal.
- Iverson, P. and P. K. Kuhl (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *The Journal of the Acoustical Society of America* 97(1), 553–562.
- Jenkins, J. (1998). The voice of the castrato. *Lancet* (351), 1877–1880.
- Johnson, K. (2011). *Acoustic and Auditory Phonetics*. Wiley-Blackwell.

-
- Kajino, S. and K. Moon (2011). Voice quality variation of Japanese porn actresses: The stylistic construction of sexual sweetness. In *Proceedings of the New Ways of Analyzing Variation 40*, Washington, DC.
- Kallio, H., R. Suviranta, M. Kuronen, and A. von Zansen (2022). Creaky voice and utterance fluency measures in predicting perceived fluency and oral proficiency of spontaneous L2 Finnish. In *Speech Prosody*, pp. 777–781. Lisbon, Portugal.
- Keating, P. (2014). Acoustic Measures of Falsetto Voice. In *Paper presented at the annual meeting of the Acoustical Society of American*. Providence, RI.
- Keating, P., C. M. Esposito, M. Garellek, S. u. D. Khan, and J. Kuang (2012). Multidimensional phonetic space for phonation contrasts. Poster presentation.
- Keating, P. and M. Garellek (2015). Acoustic analysis of creaky voice. Poster presentation. Portland, OR.
- Keating, P., M. Garellek, S. u. D. Khan, and K. Jianjing (2010). Phonation Contrasts Across Languages. *UCLA Working Papers in Phonetics* (108), 188–202.
- Keating, P., M. Garellek, and J. Kreiman (2015). Acoustic properties of different kinds of creaky voice. In *Proceedings of ICPHS XVIII*, pp. 1–5. Glasgow, UK.
- Kelman, A. (1981). Vibratory Pattern of the Vocal Folds. *Folia Phoniatrica et Logopaedica* 33(2), 73–99.
- Klatt, D. H. and L. C. Klatt (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America* 87(2), 820–857.
- Kluyver, T., B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, and C. Willing (2016). Jupyter notebooks – a publishing format for reproducible computational workflows. In F. Loizides and B. Schmidt (Eds.), *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pp. 87 – 90. IOS Press.

- Krauss, R. M., W. Apple, and L. A. Streeter (1979). Effects of Pitch and Speech Rate on Personal Attributions. *Journal of Personality and Social Psychology* 37(5), 715–727.
- Kreiman, J. (1982). Perception of sentence and paragraph boundaries in natural conversation. *Journal of Phonetics* 10(2), 163–175.
- Kreiman, J., B. R. Gerratt, M. Garellek, R. Samlan, and Z. Zhang (2014). Toward a unified theory of voice production and perception. *Loquens* 1(1), 1–9.
- Kreiman, J., B. R. Gerratt, and S. u. D. Khan (2010). Effects of native language on perception of voice quality. *Journal of Phonetics* (38), 588–593.
- Kreiman, J., B. R. Gerratt, and K. Precoda (1990). Listener Experience and Perception of Voice Quality. *Journal of Speech, Language, and Hearing Research* 33(1), 103–115.
- Kreiman, J., Y. Lee, M. Garellek, R. Samlan, and B. R. Gerratt (2021). Validating a psychoacoustic model of voice quality. *The Journal of the Acoustical Society of America* 149(1), 457–465.
- Kreiman, J., Y.-L. Shue, G. Chen, M. Iseli, B. R. Gerratt, J. Neubauer, and A. Alwan (2012). Variability in the relationships among voice quality, harmonic amplitudes, open quotient, and glottal waveform shape in sustained phonation. *The Journal of the Acoustical Society of America* 132(4), 2625–2632.
- Kreiman, J., D. Vanlancker-Sidtis, and B. R. Gerratt (2008). Perception of Voice Quality. In D. B. Pisoni and R. E. Remez (Eds.), *The Handbook of Speech Perception*, pp. 338–362. Blackwell.
- Kuang, J. (2011). Production and Perception of the Phonation Contrast in Yi. Master’s thesis, UCLA, Los Angeles.
- Ladefoged, P. (1971). *Preliminaries to linguistic phonetics*. Chicago: University of Chicago Press.
- Ladefoged, P. (2006). *A Course in Phonetics (Fifth Edition)*. Boston, USA: Thomson Wadsworth.
- Ladefoged, P. and K. Johnson (2015). *A Course in Phonetics (7th Edition)*. Australia: Wadsworth.

-
- Ladefoged, P. and I. Maddieson (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.
- Lambert, W. E., H. Frankle, and G. R. Tucker (1966). Judging personality through speech: A French-Canadian example. *16*(4), 305–321.
- Laukkanen, A.-M., E. Vilkman, P. Alku, and H. Oksanen (1996). Physical variations related to stress and emotional state: a preliminary study. *Journal of Phonetics* *24*(3), 313–335.
- Laver, J. (1980). *The Phonetic Description of Voice Quality*. Englewood Cliffs, NJ: Cambridge University Press.
- Lee, K. E. (2016). The Perception of Creaky Voice: Does Speaker Gender Affect our Judgments? Master's thesis, University of Kentucky, Lexington.
- Lee, Y., L. Goldstein, B. Parrell, and D. Byrd (2021). Who converges? Variation reveals individual speaker adaptability. *Speech Communication* (131), 23–34.
- Lee, Y., P. Keating, and J. Kreiman (2019). Acoustic voice variation within and between speakers. *The Journal of the Acoustical Society of America* *146*(3), 1568–1579.
- Lefkowitz, D. and M. A. Sicoli (2007). Creaky Voice: Constructions of gender and authority in American English conversation. In *Paper presented at the 106th annual meeting of the American Anthropological Association*, Washington, DC.
- Lehiste, I. (1965). Juncture. In *Proceedings of the 5th International Congress of Phonetic Sciences*, pp. 172–200. Münster, Germany.
- Lewandowski, E. M. and L. C. Nygaard (2018). Vocal alignment to native and non-native speakers of English. *The Journal of the Acoustical Society of America* *144*(2), 620–633.
- Lewandowski, N. and M. Jilka (2019). Phonetic Convergence, Language Talent, Personality and Attention. *Frontiers in Communication* *4*(18), 1–19.
- Lieberman, A. M. and I. G. Mattingly (1985). The motor theory of speech perception revised. *Cognition* *21*(1), 1–36.

- Lieberman, A. M. and D. H. Whalen (2000). On the relation of speech to language. *Trends in Cognitive Sciences* (4), 187–196.
- Lozano, C., A. Díaz-Negrillo, and M. Callies (2020). Designing and compiling a learner corpus of written and spoken narratives: COREFL. In C. Bongartz and J. Torregrossa (Eds.), *What's in a Narrative? Variation in Story-Telling at the Interface between Language and Literacy*, pp. 21–46. Peter Lang.
- Mendoza-Denton, N. (2007). Creaky voice in gang girl narratives. In *Paper presented at the Annual Meetings of the American Anthropological Association*, Washington, DC.
- Michaud, A. (2004). A measurement from electroglottography: DECPA, and its application in prosody. In *Proceedings of Speech Prosody*, pp. 633–636. Nara, Japan.
- Murphy, M. (2014). Sociophonetic Convergence in Native and Non-Native Speakers of French. *Concordia Working Papers in Applied Linguistics* (5), 435–450.
- Murphy, P. J., K. G. McGuigan, M. Walsh, and M. Colreavy (2008). Investigation of a glottal related harmonics-to-noise ratio and spectral tilt as indicators of glottal noise in synthesized and human voice signals. *The Journal of the Acoustical Society of America* 123(3), 1642–1652.
- Ní Chasaide, A. and C. Gobl (1997). Voice Source Variation. In W. Hardcastle and J. Laver (Eds.), *The Handbook of Phonetic Sciences*, pp. 427–461. Oxford: Blackwell Publishers Ltd.
- Ní Chasaide, A., C. Gobl, W. J. Hardcastle, and J. M. Beck (2004). Phonatory quality and affect. In W. Hardcastle and J. Beck (Eds.), *A figure of speech: A Festschrift for John Laver*, pp. 323–347. Mahwah, NJ: Routledge.
- Namy, L. L., L. C. Nygaard, and D. Sauerteig (2002). Gender Differences in Vocal Accommodation: The Role of Perception. *Journal of Language and Social Psychology* 21(4), 422–432.
- Neufeld, G. (1987). On the acquisition of prosodic and articulatory features in adult language learning. In G. Ioup and S. Weinberger (Eds.), *Interlanguage Phonology*, pp. 321–332. Cambridge, MA: Newbury House.

-
- Nilsenová, M. and M. Swerts (2012). Prosodic Adaptation in Language Learning. In J. Romero-Trillo (Ed.), *Pragmatics and Prosody in English Language Teaching*, Educational Linguistics 15, pp. 77–94. Dordrecht: Springer Science+Business Media.
- Ohara, Y. (1999). Performing gender through voice pitch: a cross-cultural analysis of Japanese American English. In *Wahrnehmung und Herstellung von Geschlecht: Perceiving and Performing Gender*, pp. 105–116. Opladen (Wiesbaden): Westdeutscher Verlag.
- Orlikoff, R. F. (1991). Assessment of the Dynamics of Vocal Fold Contact From the Electroglottogram: Data From Normal Male Subjects. *Journal of Speech, Language, and Hearing Research* 34(5), 1066–1072.
- Panfili, L. M. (2015). Linking Vowel Height and Creaky Voice.
- Papcun, G., J. Kreiman, and A. Davis (1989). Long-term memory for unfamiliar voices. *The Journal of the Acoustical Society of America* 85(2), 913–925.
- Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America* 119(4), 2382–2393.
- Pardo, J. S., I. Cajori Jay, and R. M. Krauss (2010). Conversational role influences speech imitation. *Attention, Perception, & Psychophysics* 72(8), 2254–2264.
- Pardo, J. S., R. Gibbons, A. Suppes, and R. M. Krauss (2012). Phonetic convergence in college roommates. *Journal of Phonetics* (40), 190–197.
- Pardo, J. S., I. C. Jay, R. Hoshino, S. M. Hasbun, C. Sowemimo-Coker, and R. M. Krauss (2013). Influence of Role-Switching on Phonetic Convergence in Conversation. *Discourse Processes* 50(4), 276–300.
- Pardo, J. S., A. Urmanche, H. Gash, J. Wiener, N. Mason, S. Wilman, K. Francis, and A. Decker (2019). The Montclair Map Task: Balance, Efficacy, and Efficiency in Conversational Interaction. *Language and Speech* 62(2), 378–398.
- Pardo, J. S., A. Urmanche, S. Wilman, and J. Wiener (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics* (79), 637–659.

- Pardo, J. S., A. Urmanche, S. Wilman, J. Wiener, N. Mason, K. Francis, and M. Ward (2018). A Comparison of Phonetic Convergence in Conversational Interaction and Speech Shadowing. *Journal of Phonetics* (69), 1–11.
- Parker, M. A. and S. A. Borrie (2017). Judgments of Intelligence and Likability of Young Adult Female Speakers of American English: The Influence of Vocal Fry and the Surrounding Acoustic-Prosodic Context. *Journal of Voice* 32(5), 538–545.
- Peirce, J., J. R. Gray, S. Simpson, M. MacAskill, R. Höchenberger, H. Sogo, E. Kastman, and J. K. Lindeløv (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* 51(1), 195–203.
- Pennock, B. (2004). The changing voice of women. In *Actas XXVIII Congreso Internacional AEDEAN*, pp. 1–8. Valencia, Spain.
- Pickering, M. J. and S. Garrod (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* (27), 169–190.
- Pierrehumbert, J. (1995). Prosodic effects on glottal allophones. In *Vocal Fold Physiology: Voice Quality and Control* (Singular Publishing Group ed.), pp. 39–60. San Diego: O. Fujimura & M. Hirano.
- Pierrehumbert, J. and D. Talkin (1992). Lenition of /h/ and glottal stop. In G. J. Docherty and D. R. Ladd (Eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, pp. 90–117. Cambridge: Cambridge University Press.
- Pillot-Loiseau, C., C. Horgues, S. Scheuer, and T. Kamiyama (2019). The evolution of creaky voice use in read speech by native-French and native-English speakers in tandem: a pilot study. *Anglophonia* (27), 1–24.
- Piske, T., I. R. MacKay, and J. E. Flege (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics* 29(2), 191–215.
- Pittam, J. (1987). Listeners' Evaluations of Voice Quality in Australian English Speakers. *Language and Speech* 30(2), 99–113.

-
- Podesva, R. J. (2013). Gender and the Social Meaning of Non-modal Phonation Type. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, pp. 427–448. UC Berkeley.
- Podesva, R. J. and P. Callier (2015). Voice Quality and Identity. *Annual Review of Applied Linguistics* (35), 173–194.
- Pépiot, E. (2013). *Voix de femmes, voix d’hommes : Différences acoustiques, identification du genre par la voix et implications psycholinguistiques chez les locuteurs anglophones et francophones*. Ph. D. thesis, Université Paris 8 - Vincennes-Saint Denis.
- Pépiot, E. (2014). Male and female speech: a study of mean f₀, f₀ range, phonation type and speech rate in Parisian French and American English speakers. In *Proceedings of Speech Prosody*, pp. 305–309. Dublin, Ireland.
- Purnell, T., W. Idsardi, and J. Baugh (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology* 18(1), 10–30.
- Puts, D. A. (2005). Mating context and menstrual phase affect women’s preferences for male voice pitch. *Evolution and Human Behavior* (26), 388–397.
- Puts, D. A., C. Hodges, R., R. A. Cárdenas, and S. J. C. Gaulin (2007). Men’s voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior* (28), 340–344.
- R Core Team (2022). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Also available as: www.r-project.org.
- Raphael, L. J., G. J. Borden, and K. S. Harris (2007). *Speech Science Primer: Physiology, Acoustics, and Perception of Speech*. Philadelphia, USA: Lippincott Williams and Wilkins.
- Redi, L. and S. Shattuck-Hufnagel (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics* (29), 407–429.
- Rojczyk, A. (2013). Phonetic imitation of L2 vowels in a rapid shadowing task. In *Proceedings of the 4th Pronunciation in second Language Learning and Teaching Conference*, pp. 66–76. Ames, IA.

- Rothenberg, M. (1979). Some relations between glottal air flow and vocal fold contact area. *ASHA Reports* (11), 88–96.
- RStudio Team (2015). *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, Inc. Also available as: www.rstudio.com.
- Sancier, M. L. and C. A. Fowler (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25(4), 421–436.
- Sapir, E. (1927). Speech as a Personality Trait. *American Journal of Sociology* (32), 892–905.
- Sato, M., K. Grabski, M. Garnier, L. Granjon, J.-L. Schwartz, and N. Nguyen (2013). Converging towards a common speech code: imitative and perceptuo-motor recalibration processes in speech production. *Frontiers in Cognitive Sciences* 4(422), 1–14.
- Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin* 99(2), 143–165.
- Shue, Y.-L. (2010). *The voice source in speech production: Data, analysis and models*. Ph. D. thesis, UCLA, Los Angeles.
- Shue, Y.-L., P. Keating, C. Vicenik, and K. Yu (2011). VoiceSauce: A Program for Voice Analysis. In *Proceedings of ICPHS XVII*, pp. 1846–1849. Hong Kong, China.
- Sicoli, M. A. (2010). Shifting voices with participant roles: Voice qualities and speech registers in Mesoamerica. *Language in Society* 39(4), 521–553.
- Silverman, D., B. Blankenship, P. Kirk, and P. Ladefoged (1995). Phonetic structures in Jalapa Mazatec. *Anthropological Linguistics* (37), 70–88.
- Simpson, A. P. (2009). Phonetic differences between male and female speech. *Language and Linguistics Compass* 3(2), 621–640.
- Slifka, J. (2000). *Respiratory constraints on speech production at prosodic boundaries*. Ph. D. thesis, MIT, Cambridge, MA.

-
- Slifka, J. (2006). Some Physiological Correlates to Regular and Irregular Phonation at the End of an Utterance. *Journal of Voice* 20(2), 171–186.
- Smith, A. M. and D. G. Childers (1983). Laryngeal Evaluation Using Features from Speech and the Electroglottograph. *IEEE Transactions on Biomedical Engineering* 30(11), 755–759.
- Smith, C. (2007). Prosodic accommodation by French speakers to a non-native interlocutor. In *Proceedings of ICPHS XVI*, pp. 1081–1084. Saarbrücken, Germany.
- Snow, C. (1995). Issues in the study of input: Fine-tuning, universality, individual and developmental differences and necessary causes. In P. Fletcher and B. MacWhinney (Eds.), *The Handbook of Child Language*, pp. 180–194. Oxford: Blackwell.
- Stel, M., R. B. Van Baaren, J. Blascovich, E. van Dijk, M. Cade, M. M. H. Pollmann, M. van Leeuwen, J. Mastop, and R. Vonk (2010). Effects of a priori liking on the elicitation of mimicry. *Experimental Psychology* 57(6), 412–418.
- Stel, M., R. B. Van Baaren, and R. Vonk (2008). Effects of mimicking: acting prosocially by being emotionally moved. *European Journal of Social Psychology* (38), 965–976.
- Stevens, K. N. (1977). Physics of Laryngeal Behavior and Larynx Modes. *Phonetica* 34(4), 264–279.
- Stuart-Smith, J. (1999). Glasgow: Accent and voice quality. In P. Foulkes and G. Doherty (Eds.), *Urban Voices: Accent Studies in the British Isles*, pp. 201–222. Leeds, UK: Arnold.
- Swerts, M. and S. Zerbian (2010). Intonational Differences between L1 and L2 English in South Africa. *Phonetica* 67(3), 127–146.
- Tehrani, H. (2009). EGGWorks: a program for automated analysis of EGG signals.
- Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America* 85(4), 1699–1707.
- Titze, I. R. (1994). *Principles of Voice Production*. Englewood Cliffs, NJ: Prentice-Hall Inc.
- Trager, G. L. (1958). Paralanguage: A First Approximation. *Studies in Linguistics* (13), 1–12.

- Umeda, N. (1978). Occurrence of glottal stops in fluent speech. *The Journal of the Acoustical Society of America* 64(1), 88–94.
- Vaissière, J. (2002). Cross-linguistic prosodic transcription: French vs. English. In *Problems and methods of experimental phonetics. In honour of the 70th anniversaire of Pr. L. V. Bondarko*, pp. 147–164. St Petersburg: St Petersburg State University Press.
- van den Berg, J. (1958). Myoelastic-aerodynamic theory of voice production. *Journal of Speech and Hearing Research* 1(3), 227–244.
- Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S (Fourth Edition)*. New York: Springer.
- Wagner, M., M. Broersma, J. M. McQueen, S. Dhaene, and K. Lemhöfer (2021). Phonetic convergence to non-native speech: Acoustic and perceptual evidence. *Journal of Phonetics* (88), 1–20.
- Wang, X. and W. Gu (2002). Effects of Gender and Language Proficiency on Phonetic Accommodation in Chinese EFL Learners. In *Proceedings of Speech Prosody*, pp. 469–772. Lisbon, Portugal.
- Waniek-Klimczak, E. (1998). Sources of variability in L2 phonology: vowel duration in 'Polish English'. In C. C. Gruber, D. Higgins, K. Olson, and T. Wycsocki (Eds.), *CLS 34/2: The Panels*, pp. 387–401. Chicago: Chicago Linguistic Society.
- Watzlawick, P., J. B. Bavelas, and D. D. Jackson (1967). *Pragmatics of Human Communication: A Study of Interactional Patterns, Pathologies and Paradoxes*. New-York: Norton.
- Wells, J. C. (2000). *Longman Pronunciation Dictionary*. London: Pearson Education Limited.
- Wendhal, R. W. (1963). Laryngeal Analog Synthesis of Harsh Voice Quality. *Folia Phoniatica et Logopaedica* (15), 241–250.
- Whalen, D. H. and A. G. Levitt (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics* (23), 349–366.

-
- Willemyns, M., C. Gallois, V. J. Callan, and J. Pittam (1997). Accent Accommodation in the Job Interview: Impact of Interviewer Accent and Gender. *Journal of Language and Social Psychology* 16(1), 3–22.
- Winter, B. (2020). *Statistics for Linguists: An Introduction Using R*. New-York: Routledge.
- Wolk, L., N. B. Abdelli-Beruh, and D. Slavin (2012). Habitual Use of Vocal Fry in Young Adult Female Speakers. *Journal of Voice* 26(3), 111–116.
- Wright, R., C. Mansfield, and L. Panfili (2019). Voice quality types and uses in North American English. *Anglophonia* (27), 1–15.
- Yamazawa, H. and H. Hollien (1992). Speaking Fundamental Frequency Patterns of Japanese Women. *Phonetica* (49), 128–140.
- Yoonjeong, L., L. Goldstein, B. Parrell, and D. Byrd (2021). Who converges? Variation reveals individual speaker adaptability. *Speech Communication* (131), 23–34.
- Yuasa, I. P. (2010). Creaky voice: A new feminine voice quality for young urban-oriented upwardly mobile American women? *American Speech* 85(3), 315–337.
- Zajac, M. (2013). Phonetic Imitation of Vowel Duration in L2 Speech. *Research in Language* 11(1), 19–29.
- Zemlin, W. R. (1998). *Speech and Hearing Science: Anatomy and Physiology (4th edition)*. Boston: Allyn and Bacon.
- Zhang, S. and N. Thompson (2004). DIALANG: A Diagnostic Language Assessment System (review). *The Canadian Modern Language Review/La revue canadienne des langues vivantes* 61(2), 290–293.
- Zhang, Z. (2016). Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model. *The Journal of the Acoustical Society of America* 139(4), 1493–1507.

Appendix A

Stimuli

A.1 List of the English sentences presented to the American model speakers

1. They had a running gag.
2. That balloon is big.
3. She became very mad.
4. He felt very sad.
5. She ate a ripe fig.
6. This TV show is bad.
7. She performed a good deed.
8. The wound formed a scab.
9. She saw a bright colored bead.
10. She was walking her dog.
11. She has a new gig.
12. My shirt is in my bag.
13. He swallowed a seed.
14. Raccoons love to dig.
15. There was too much fog.
16. He painted a nude.
17. He's waiting for his dad.
18. The meadow was full of bees.
19. I'm not in the mood.

20. She poked him with a stab.
21. It was hard to seize.
22. The farmer bought more feed.
23. I've had too much food.
24. He bought a bottle of booze.
25. He sailed the seven seas.
26. It moved at a great speed.
27. The worker carried some saws.

A.2 List of the French sentences presented to the French model speakers

1. Il descendit à la cave.
2. Il détestait les figues.
3. Elle marcha dans une bouse.
4. Il avait mal au bide.
5. Elle avait fait une fugue.
6. Ils retournaient à la base.
7. Il avait peur du vide.
8. Elle portait une belle bague.
9. Ils se faisaient la bise.
10. C'était devenu sa muse.
11. Il fallait qu'il bouge.

12. Il possédait trois dagues.
13. On aperçut une buse.
14. Elle marchait sur la digue.
15. Il y mit de la soude.
16. Elle était pleine de bave.
17. Il lui offrit un vase.
18. Il criait dans le stade.
19. Elle traversa à la nage.
20. Il en avait douze.
21. Ca sentait fort le gaz.
22. La soupe était trop fade.
23. Sa mère était guide.
24. Sa réponse était trop vague.
25. Ils lui donnèrent un gage.
26. Les couleurs étaient vives.
27. Elle était vraiment sage.

A.3 Phonological transcription of English and French sentence-final words

Transcription of sentence-final words only is given, for they are the focus of this study. Transcription of English words are given in General American (Wells, 2000).

Table A.1: Phonological transcription of sentence-final words.

| Sentence | English | French |
|----------|---------|--------|
| 1 | /gæg/ | /kav/ |
| 2 | /bɪg/ | /fɪg/ |
| 3 | /mæd/ | /buz/ |
| 4 | /sæd/ | /bid/ |
| 5 | /fɪg/ | /fyg/ |
| 6 | /bæd/ | /baz/ |
| 7 | /di:d/ | /vid/ |
| 8 | /skæb/ | /bæg/ |
| 9 | /bi:d/ | /biz/ |
| 10 | /dæg/ | /myz/ |
| 11 | /gɪg/ | /buʒ/ |
| 12 | /bæg/ | /dag/ |
| 13 | /si:d/ | /byz/ |
| 14 | /dɪg/ | /dig/ |
| 15 | /fæg/ | /sud/ |
| 16 | /nu:d/ | /bav/ |
| 17 | /dæg/ | /vaz/ |
| 18 | /bi:z/ | /stad/ |
| 19 | /mu:d/ | /naʒ/ |
| 20 | /stæb/ | /duz/ |
| 21 | /si:z/ | /gaz/ |
| 22 | /fi:d/ | /fad/ |
| 23 | /fu:d/ | /gid/ |
| 24 | /bu:z/ | /vag/ |
| 25 | /si:z/ | /gaʒ/ |
| 26 | /spi:d/ | /viv/ |
| 27 | /saz/ | /saʒ/ |

A.4 Key to phonetic/SAMPA symbols for English and French

For English, only symbols present in the English dialect called 'General American' are indicated for this dialect is used as the reference accent in this study. The reference used for each phoneme corresponds to the one given in Wells' Lexical Set. These references are also those used in the analysis of vowels.

Table A.2: American English consonants.

| General American | SAMPA | Reference |
|------------------|-------|-------------------|
| p | p | <u>p</u> en |
| b | b | <u>b</u> ack |
| t | t | <u>t</u> ea |
| d | d | <u>d</u> ay |
| k | k | <u>k</u> ey |
| g | g | <u>g</u> et |
| tʃ | tS | <u>ch</u> urch |
| dʒ | dZ | <u>j</u> udge |
| f | f | <u>f</u> at |
| v | v | <u>v</u> iew |
| θ | T | <u>t</u> hing |
| ð | D | <u>t</u> his |
| s | s | <u>s</u> oon |
| z | z | <u>z</u> ero |
| ʃ | S | <u>sh</u> ip |
| ʒ | Z | <u>pl</u> easeure |
| h | h | <u>h</u> ot |
| m | m | <u>m</u> ore |
| n | n | <u>n</u> ice |
| ŋ | N | <u>r</u> ing |
| l | l | <u>l</u> ight |
| r | r | <u>r</u> ight |
| j | j | <u>y</u> et |
| w | w | <u>w</u> et |

Table A.3: American English vowels.

| General American | SAMPA | Reference |
|------------------|-------|------------|
| ɪ | I | kit |
| ɛ | E | dress |
| æ | { | trap |
| ʌ | V | strut |
| ʊ | U | foot |
| i | i | fleece |
| eɪ | eI | face |
| aɪ | aI | price |
| ɔɪ | OI | choice |
| u | u | goose |
| oʊ | oU | goat |
| aʊ | aU | mouth |
| ɑɪ | Ar | start, lot |
| ɔ | O | north |
| ɝ˞ | ɜ' | nurse |
| ə | @ | about |

Table A.4: French consonants.

| French | SAMPA | Reference |
|--------|-------|------------------|
| p | p | <u>p</u> ont |
| b | b | <u>b</u> on |
| t | t | <u>t</u> emps |
| d | d | <u>d</u> ans |
| k | k | <u>q</u> uand |
| g | g | <u>g</u> ant |
| f | f | <u>f</u> emme |
| v | v | <u>v</u> ent |
| s | s | <u>s</u> ans |
| z | z | <u>z</u> one |
| ʃ | S | <u>ch</u> amps |
| ʒ | Z | <u>g</u> ens |
| j | j | <u>i</u> on |
| m | m | <u>m</u> ont |
| n | n | <u>n</u> om |
| ɲ | n | oi <u>gn</u> on |
| ŋ | N | ca <u>mp</u> ing |
| l | l | <u>l</u> ong |
| r | R | <u>r</u> ond |
| w | w | co <u>i</u> n |
| ɥ | H | ju <u>i</u> n |

Table A.5: French vowels.

| General American | SAMPA | Reference |
|------------------|-------|-----------------------------|
| i | i | s <u>i</u> |
| e | e | se <u>s</u> |
| ɛ | E | se <u>i</u> ze |
| a | a | pa <u>t</u> te |
| ɑ | A | pa <u>t</u> te |
| ɔ | O | co <u>m</u> me |
| o | o | gr <u>o</u> s |
| u | u | do <u>u</u> x |
| y | y | du <u>u</u> |
| ø | 2 | de <u>u</u> x |
| œ | 9 | ne <u>u</u> f |
| ə | @ | ju <u>s</u> t <u>e</u> ment |
| ẽ | e~ | vi <u>n</u> |
| ã | a~ | ve <u>n</u> t |
| õ | o~ | bo <u>n</u> |
| õe | 9~ | bru <u>n</u> |

A.5 List of the stimuli selected for each model talker

A.5. List of the stimuli selected for each model talker

Table A.6: List of the stimuli selected for each model speaker.

| Sentence | AM_01 | AM_02 | AM_03 | AF_01 | AF_04 | AF_06 | FM_02 | FM_05 | FM_06 | FF_01 | FF_02 | FF_08 |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1 | | | X | X | X | X | | | X | | X | |
| 2 | X | | X | X | X | | | | | | | |
| 3 | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | |
| 5 | X | | X | X | X | | | | | | | |
| 6 | | | | X | X | X | | X | X | X | | X |
| 7 | | | | X | X | X | | | | | | |
| 8 | | | | | | | | | | | X | |
| 9 | | | | | | | | X | | X | | X |
| 10 | | | | | | | | X | | | | |
| 11 | X | | X | X | X | | | X | X | | X | X |
| 12 | | | | | | | | | | | | |
| 13 | | X | | X | X | | X | | | X | | |
| 14 | X | | X | X | | X | | X | | X | | X |
| 15 | | | | | | | X | | X | | | X |
| 16 | X | X | | X | | | | X | X | X | | X |
| 17 | | | | | | | X | X | X | | X | X |
| 18 | | | | | | | X | | | | X | |
| 19 | X | X | X | | | | | X | X | | X | X |
| 20 | | | | | | | | | | | | |
| 21 | X | X | | X | | | | | | | | |
| 22 | X | X | | X | | | X | | X | X | X | |
| 23 | | X | X | | | | X | | X | X | X | X |
| 24 | | X | X | | X | | | | | | | |
| 25 | | X | | X | X | | | X | X | | X | X |
| 26 | | X | X | | | X | X | | | X | | |
| 27 | | X | X | | | | X | X | X | X | X | |

Appendix B

Analysis

- B.1 Convergence and divergence effects for each measure depending on model speaker

Table B.1: Convergence and divergence effects for each measure depending on model speaker. **C** is indicated when convergence was observed, **D** when divergence was observed.

| MS | AF_01 | AF_04 | AF_06 | AM_01 | AM_02 | AM_03 | FF_01 | FF_02 | FF_08 | FM_02 | FM_05 | FM_06 |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Duration | C | D | D | C | C | D | C | C | C | D | C | D |
| f_0 | C | D | D | C | C | C | D | D | D | C | C | D |
| H1* | D | D | C | D | D | D | D | D | D | D | D | D |
| H2* | D | D | D | D | D | D | C | D | D | D | D | D |
| H1*-A1* | D | D | C | C | D | D | C | D | D | D | D | D |
| H1*-A2* | C | D | D | D | D | D | D | D | D | D | D | D |
| H1*-A3* | C | C | D | D | D | D | D | D | D | D | D | D |
| H2*-H4* | C | D | D | C | D | C | C | C | C | C | D | C |
| CPP | D | C | D | D | D | D | D | C | C | C | C | D |
| CQ | D | D | C | C | D | C | C | C | C | D | D | D |
| PIC | D | C | D | C | C | C | C | C | C | D | D | D |

B.2 Linear mixed-effects models

Best linear mixed-effects models found for each acoustic and EGG measure in the accommodation analysis.

Table B.2: Best model found for duration DID score (Conditional $R^2 = 0.072$).

| <i>AIC</i> | <i>BIC</i> | <i>LogLik</i> | <i>Deviance</i> | <i>df.resid</i> | |
|------------------------|-----------------|-----------------|-----------------|-----------------|-------------|
| 22797.5 | 22820.4 | -11394.8 | 22789.5 | 2226 | |
| Random effects: | | | | | |
| <i>Groups</i> | <i>Name</i> | <i>Variance</i> | <i>SD</i> | | |
| Speaker | (Intercept) | 27.47 | 5.241 | | |
| MS | (Intercept) | 94.85 | 9.739 | | |
| Residual | | 1570.11 | 39.625 | | |
| Fixed effects: | | | | | |
| | <i>Estimate</i> | <i>SE</i> | <i>df</i> | <i>Z</i> | <i>p(Z)</i> |
| (Intercept) | 4.067 | 3.160 | 15.429 | 1.287 | 0.217 |

Table B.3: Best model found for f_0 DID score (Conditional $R^2 = 0.039$).

| <i>AIC</i> | <i>BIC</i> | <i>LogLik</i> | <i>Deviance</i> | <i>df.resid</i> | |
|--------------------------------------|-----------------|-----------------|-----------------|-----------------|--------------|
| 22569.8 | 22604.0 | -11278.9 | 22557.8 | 2224 | |
| Random effects: | | | | | |
| <i>Groups</i> | <i>Name</i> | <i>Variance</i> | <i>SD</i> | | |
| Speaker | (Intercept) | 39.88 | 6.315 | | |
| Residual | | 1429.43 | 37.808 | | |
| Fixed effects: | | | | | |
| | <i>Estimate</i> | <i>SE</i> | <i>df</i> | <i>Z</i> | <i>p(Z)</i> |
| (Intercept) | 0.2115 | 2.1018 | 56.1276 | 0.101 | 0.920214 |
| Language:French | -3.1787 | 2.2620 | 2210.6654 | -1.405 | 0.160091 |
| Gender:Male | 8.6558 | 2.2349 | 2210.3789 | 3.873 | 0.000111 *** |
| Lngg:Fr×Gndr:Male | -4.1731 | 3.2050 | 2210.2776 | -1.302 | 0.193031 |
| Correlation of Fixed Effects: | | | | | |
| | (Intercept) | Lngg:Fr | Gndr:Male | | |
| Lngg:Fr | -0.510 | | | | |
| Gndr:Male | -0.516 | 0.479 | | | |
| Lngg:Frh×Gndr:Male | 0.360 | -0.706 | -0.697 | | |

Table B.4: Best model found for H1* DID score (Conditional $R^2 = 0.228$).

| <i>AIC</i> | <i>BIC</i> | <i>LogLik</i> | <i>Deviance</i> | <i>df.resid</i> | |
|------------------------|-----------------|-----------------|-----------------|-----------------|-------------|
| 13662.7 | 13685.6 | -6827.4 | 13654.7 | 2226 | |
| Random effects: | | | | | |
| <i>Groups</i> | <i>Name</i> | <i>Variance</i> | <i>SD</i> | | |
| Speaker | (Intercept) | 5.009 | 2.238 | | |
| MS | (Intercept) | 2.556 | 1.599 | | |
| Residual | | 25.594 | 5.059 | | |
| Fixed effects: | | | | | |
| | <i>Estimate</i> | <i>SE</i> | <i>df</i> | <i>Z</i> | <i>p(Z)</i> |
| (Intercept) | -1.7901 | 0.6892 | 28.6373 | -2.597 | 0.0147 * |

Table B.5: Best model found for H2* DID score (Conditional $R^2 = 0.179$).

| <i>AIC</i> | <i>BIC</i> | <i>LogLik</i> | <i>Deviance</i> | <i>df.resid</i> | |
|------------------------|-----------------|-----------------|-----------------|-----------------|-------------|
| 14587.6 | 14610.5 | -7289.8 | 14579.6 | 2226 | |
| Random effects: | | | | | |
| <i>Groups</i> | <i>Name</i> | <i>Variance</i> | <i>SD</i> | | |
| Speaker | (Intercept) | 6.126 | 2.475 | | |
| MS | (Intercept) | 2.373 | 1.540 | | |
| Residual | | 38.912 | 6.238 | | |
| Fixed effects: | | | | | |
| | <i>Estimate</i> | <i>SE</i> | <i>df</i> | <i>Z</i> | <i>p(Z)</i> |
| (Intercept) | -1.5163 | 0.7222 | 28.9311 | -2.099 | 0.0446 * |

Table B.6: Best model found for H1*-H2* DID score (Conditional $R^2 = 0.018$).

| <i>AIC</i> | <i>BIC</i> | <i>LogLik</i> | <i>Deviance</i> | <i>df.resid</i> | |
|------------------------|-----------------|-----------------|-----------------|-----------------|-------------|
| 13695.9 | 13718.8 | -6844.0 | 13687.9 | 2226 | |
| Random effects: | | | | | |
| <i>Groups</i> | <i>Name</i> | <i>Variance</i> | <i>SD</i> | | |
| Speaker | (Intercept) | 0.1854 | 0.4306 | | |
| MS | (Intercept) | 0.3035 | 0.5509 | | |
| Residual | | 26.8201 | 5.1788 | | |
| Fixed effects: | | | | | |
| | <i>Estimate</i> | <i>SE</i> | <i>df</i> | <i>Z</i> | <i>p(Z)</i> |
| (Intercept) | -0.2603 | 0.2160 | 14.9127 | -1.205 | 0.247 |

Table B.7: Best model found for H1*-A1* DID score (Conditional $R^2 = 0.015$).

| <i>AIC</i> | <i>BIC</i> | <i>LogLik</i> | <i>Deviance</i> | <i>df.resid</i> | |
|------------------------|-----------------|-----------------|-----------------|-----------------|-------------|
| 14457.0 | 14479.8 | -7224.5 | 14449.0 | 2226 | |
| Random effects: | | | | | |
| <i>Groups</i> | <i>Name</i> | <i>Variance</i> | <i>SD</i> | | |
| Speaker | (Intercept) | 0.4388 | 0.6624 | | |
| MS | (Intercept) | 0.1352 | 0.3678 | | |
| Residual | | 37.7660 | 6.1454 | | |
| Fixed effects: | | | | | |
| | <i>Estimate</i> | <i>SE</i> | <i>df</i> | <i>Z</i> | <i>p(Z)</i> |
| (Intercept) | -0.5038 | 0.2241 | 17.5228 | -2.248 | 0.0377 * |

Table B.8: Best model found for H1*-A2* DID score (Conditional $R^2 = 0.027$).

| <i>AIC</i> | <i>BIC</i> | <i>LogLik</i> | <i>Deviance</i> | <i>df.resid</i> | |
|------------------------|-----------------|-----------------|-----------------|-----------------|-------------|
| 14691.8 | 14714.7 | -7341.9 | 14683.8 | 2226 | |
| Random effects: | | | | | |
| <i>Groups</i> | <i>Name</i> | <i>Variance</i> | <i>SD</i> | | |
| Speaker | (Intercept) | 0.3595 | 0.5996 | | |
| MS | (Intercept) | 0.4653 | 0.6821 | | |
| Residual | | 41.8853 | 6.4719 | | |
| Fixed effects: | | | | | |
| | <i>Estimate</i> | <i>SE</i> | <i>df</i> | <i>Z</i> | <i>p(Z)</i> |
| (Intercept) | -0.6087 | 0.2750 | 16.5522 | -2.213 | 0.0412 * |

Table B.9: Best model found for H1*-A3* DID score (Conditional $R^2 = 0.098$).

| <i>AIC</i> | <i>BIC</i> | <i>LogLik</i> | <i>Deviance</i> | <i>df.resid</i> | |
|------------------------|-----------------|-----------------|-----------------|-----------------|-------------|
| 14881.3 | 14904.1 | -7436.7 | 14873.3 | 2226 | |
| Random effects: | | | | | |
| <i>Groups</i> | <i>Name</i> | <i>Variance</i> | <i>SD</i> | | |
| Speaker | (Intercept) | 0.5803 | 0.7618 | | |
| MS | (Intercept) | 1.0553 | 1.0273 | | |
| Residual | | 45.3827 | 6.7367 | | |
| Fixed effects: | | | | | |
| | <i>Estimate</i> | <i>SE</i> | <i>df</i> | <i>Z</i> | <i>p(Z)</i> |
| (Intercept) | -0.6133 | 0.3708 | 17.1602 | -1.654 | 0.116 |

Table B.10: Best model found for H2*-H4* DID score (Conditional $R^2 = 0.013$).

| <i>AIC</i> | <i>BIC</i> | <i>LogLik</i> | <i>Deviance</i> | <i>df.resid</i> | |
|------------------------|-----------------|-----------------|-----------------|-----------------|-------------|
| 14337.1 | 14359.9 | -7164.5 | 14329.1 | 2226 | |
| Random effects: | | | | | |
| <i>Groups</i> | <i>Name</i> | <i>Variance</i> | <i>SD</i> | | |
| Speaker | (Intercept) | 0.2846 | 0.5335 | | |
| MS | (Intercept) | 0.1969 | 0.4437 | | |
| Residual | | 35.8150 | 5.9846 | | |
| Fixed effects: | | | | | |
| | <i>Estimate</i> | <i>SE</i> | <i>df</i> | <i>Z</i> | <i>p(Z)</i> |
| (Intercept) | -0.04076 | 0.21626 | 16.12239 | -0.188 | 0.853 |

Table B.11: Best model found for CPP DID score (Conditional $R^2 = 0.073$).

| <i>AIC</i> | <i>BIC</i> | <i>LogLik</i> | <i>Deviance</i> | <i>df.resid</i> | |
|------------------------|-----------------|-----------------|-----------------|-----------------|-------------|
| 11059.3 | 11082.1 | -5525.6 | 11051.3 | 2226 | |
| Random effects: | | | | | |
| <i>Groups</i> | <i>Name</i> | <i>Variance</i> | <i>SD</i> | | |
| Speaker | (Intercept) | 0.1395 | 0.3735 | | |
| MS | (Intercept) | 0.5006 | 0.7076 | | |
| Residual | | 8.1264 | 2.8507 | | |
| Fixed effects: | | | | | |
| | <i>Estimate</i> | <i>SE</i> | <i>df</i> | <i>Z</i> | <i>p(Z)</i> |
| (Intercept) | -0.002486 | 0.228847 | 15.354539 | -0.011 | 0.991 |

Table B.12: Best model found for CQ DID score (Conditional $R^2 = 0.077$).

| <i>AIC</i> | <i>BIC</i> | <i>LogLik</i> | <i>Deviance</i> | <i>df.resid</i> | |
|------------------------|-----------------|-----------------|-----------------|-----------------|-------------|
| -2496.9 | -2474.1 | 1252.5 | -2504.9 | 2226 | |
| Random effects: | | | | | |
| <i>Groups</i> | <i>Name</i> | <i>Variance</i> | <i>SD</i> | | |
| Speaker | (Intercept) | 0.0014375 | 0.03791 | | |
| MS | (Intercept) | 0.0001175 | 0.01084 | | |
| Residual | | 0.0185862 | 0.13633 | | |
| Fixed effects: | | | | | |
| | <i>Estimate</i> | <i>SE</i> | <i>df</i> | <i>Z</i> | <i>p(Z)</i> |
| (Intercept) | -0.002746 | 0.009489 | 23.037118 | -0.289 | 0.775 |

Table B.13: Best model found for PIC DID score (Conditional $R^2 = 0.13$).

| <i>AIC</i> | <i>BIC</i> | <i>LogLik</i> | <i>Deviance</i> | <i>df.resid</i> | |
|------------------------|-----------------|-----------------|-----------------|-----------------|-------------|
| 24348.2 | 24371.0 | -12170.1 | 24340.2 | 2226 | |
| Random effects: | | | | | |
| <i>Groups</i> | <i>Name</i> | <i>Variance</i> | <i>SD</i> | | |
| Speaker | (Intercept) | 396.99 | 19.925 | | |
| MS | (Intercept) | 67.65 | 8.225 | | |
| Residual | | 3116.67 | 55.827 | | |
| Fixed effects: | | | | | |
| | <i>Estimate</i> | <i>SE</i> | <i>df</i> | <i>Z</i> | <i>p(Z)</i> |
| (Intercept) | 1.024 | 5.186 | 26.575 | 0.197 | 0.845 |

Appendix C

Metadata

C.1 Model speakers

This reading task will be conducted with an electroglottograph (EGG). Electroglottography is a non-invasive, electrophysiological technique, that allows the observation of the properties of the vocal folds during voiced sounds. Two electrodes are positioned on the subject's neck over the thyroid cartilages. This technique does not represent any risk for health. The length of the experiment will not exceed 30 minutes.

Cette tâche de lecture sera conduite avec un électroglottographe (EGG). L'électroglottographie est une technique de mesure de la surface d'accolement des cordes vocales au cours de la phonation. Cette technique, non-invasive, consiste à placer un collier muni de deux électrodes de part et d'autre du larynx. Elle ne présente aucun danger particulier, ni contre-indication connue. La durée de la tâche n'excédera pas 30 minutes.

QUESTIONNAIRE

Age / Age :

Sex / Sexe :

Gender / Genre :

Place of birth / Lieu de naissance :

Native language(s) / Langue(s) maternelle(s) :

Mother's native language(s) / Langue(s) maternelle(s) de la mère :

Father's native language(s) / Langue(s) maternelle(s) du père :

Place(s) of residence (until 10) / Lieu(x) de résidence (jusqu'à 10 ans) :

Other(s) (after 10) / Autres (après 10 ans) :

Length / Durée :

Occupation / Occupation :

Level of education / Niveau d'études :

Smoker—Non-smoker / Fumeur.se—Non-fumeur.se :

Other spoken language(s) / Autre(s) langue(s) parlée(s) :

Level / Niveau :

History of speech/hearing disorder (if yes, specify) / Antécédents de trouble de la voix ou de l'audition ? (si oui, spécifiez lesquels) :

Table C.1: Male American English models' metadata.

| Model speaker | <i>AM_01</i> | <i>AM_02</i> | <i>AM_03</i> |
|---|------------------------|---------------------|--------------------------------------|
| Age | 26 | 24 | 25 |
| Sex | M | M | M |
| Gender | M | M | M |
| Birthplace | Seattle, WA | Seattle, WA | San Ramon, CA |
| Native language(s) | English | English | English |
| Mother's native language(s) | English | English | English |
| Father's native language(s) | English | English | French |
| Place(s) of residence (until 10) | Seattle, WA | Seattle, WA | California Maryland Washington |
| Other(s) (after 10) | Madrid Morocco | Ravensdale, WA | Washington |
| Length | 1 year 1 year | 22 years | 22 years |
| Occupation | Student | Manager | Bakery Manager |
| Level of education | MS | High School Diploma | High School Diploma |
| Smoker/Non-smoker | Non-smoker | Non-smoker | Smoker |
| Other spoken languages | French Spanish | | French |
| Level | Fluent Intermediate | | Intermediate |
| History of speech/hearing disorder | No | No | No |

Table C.2: Female American English models' metadata.

| Model speaker | <i>AF_01</i> | <i>AF_04</i> | <i>AF_06</i> |
|------------------------------------|--|---|---|
| Age | 28 | 25 | 32 |
| Sex | F | F | F |
| Gender | F | F | F |
| Birthplace | Los Angeles, CA | California | Fremont, CA |
| Native language(s) | English/Spanish | English | English |
| Mother's native language(s) | Korean | English | English |
| Father's native language(s) | English | Korean | English |
| Place(s) of residence (until 10) | Los Angeles, CA | Palo Alto, CA | Union City, CA |
| Other(s) (after 10) | Santa Cruz, CA Barcelona Boston, MA Seattle, WA | Boston, MA Pittsburgh, PA | California The Netherlands France |
| Length | 5 years 1 year 3 years 3 years | 4 years 2 years | 31 years 5 months 8 months |
| Occupation | PhD student | PhD student | Speech therapist |
| Level of education | MA | MS | MS |
| Smoker/Non-smoker | Non-smoker | Non-smoker | Non-smoker |
| Other spoken languages | Spanish Korean Russian French | Mandarin Korean Spanish Russian | French Spanish Dutch |
| Level | Fluent Intermediate Intermediate Intermediate | Intermediate Near-native Advanced Beginner | Intermediate Beginner Beginner |
| History of speech/hearing disorder | No | No | No |

Table C.3: Male French models' metadata.

| Model speaker | <i>FM_02</i> | <i>FM_05</i> | <i>FM_06</i> |
|---|--------------|--------------------------|----------------------------------|
| Age | 27 | 30 | 23 |
| Sex | M | M | M |
| Gender | M | M | M |
| Birthplace | Paris | Casablanca | Lille |
| Native language | French | French | French |
| Mother's native language(s) | French | French | French |
| Father's native language(s) | French | French | French |
| Place(s) of residence (until 10) | France | Morocco | France |
| Place(s) of residence (after 10) | France | France | France |
| Length | 17 years | 12 years | 17 years |
| Occupation | PhD student | Restaurant manager | Student |
| Level of education | MA | MA | BA |
| Smoker/Non-smoker | Smoker | Non-smoker | Non-smoker |
| Other spoken languages | English | English Spanish | German English Italian |
| Level | Advanced | Advanced Intermediate | Advanced Advanced Beginner |
| History of speech/hearing disorder | No | No | No |

Table C.4: Female French models' metadata.

| Model speaker | <i>FF_01</i> | <i>FF_02</i> | <i>FF_08</i> |
|---|--------------------------|------------------------------|--------------|
| Age | 27 | 28 | 23 |
| Sex | F | F | F |
| Gender | F | F | F |
| Birthplace | Laval | Troyes | Versailles |
| Native language | French | French | French |
| Mother's native language(s) | French | Arabic | French |
| Father's native language(s) | French | Arabic | French |
| Place(s) of residence (until 10) | France | Morocco | France |
| Place(s) of residence (after 10) | France | France France Vietnam | France |
| Length | 17 years | 25 years 3 years | 13 years |
| Occupation | Student | Head waiter | Student |
| Level of education | BA | High School Diploma | BA |
| Smoker/Non-smoker | Non-smoker | Smoker (occasionally) | Non-smoker |
| Other spoken languages | English German | English Arabic | English |
| Level | Advanced Intermediate | Intermediate Intermediate | Advanced |
| History of speech/hearing disorder | No | No | No |

C.2 Participants

Table C.5: Number of hours/week spent reading, writing, speaking, and listening to English reported by each subject. Subjects also specified whether they would listen more often to British or American English varieties.

| Subject | Reading | Writing | Speaking | Listening (British/American) |
|----------------|----------------|----------------|-----------------|-------------------------------------|
| <i>S1</i> | 9 | 5 | 1 | 11 (American) |
| <i>S2</i> | 14 | 8 | 6 | 3 (American) |
| <i>S3</i> | 12 | 12 | 10 | 5 (British) |
| <i>S4</i> | 20 | 20 | 30 | 40 (American) |
| <i>S6</i> | 5 | 1 | 0 | 15 (American) |
| <i>S7</i> | 3 | 6 | 1 | 12 (American) |
| <i>S8</i> | NA | 6 | 2 | 10 (British) |
| <i>S10</i> | 14 | 20 | 10 | 5 (American) |
| <i>S13</i> | 14 | 12 | 3 | 7 (American) |
| <i>S14</i> | 3 | 6 | 2 | 10 (American) |
| <i>S15</i> | 10 | 10 | 10 | 10 (American) |
| <i>S16</i> | 6 | 9 | 4 | 15 (American) |
| <i>S17</i> | 8 | 12 | 5 | 10 (American) |
| <i>S20</i> | 8 | 20 | 10 | 15 (American) |
| <i>S22</i> | 7 | 3 | 1 | 30 (American) |
| <i>S23</i> | 14 | 5 | 2 | 50 (American) |
| <i>S24</i> | 2 | 8 | 1 | 16 (American) |
| <i>S27</i> | 10 | 6 | 5 | 10 (American) |
| <i>S28</i> | 4 | 3 | 6 | 6 (American) |
| <i>S29</i> | 10 | 6 | 1 | 37 (American) |

Table C.6: Each subject reported the amount of time they spent in an English-speaking country in months (1 week = 0.25 months). They self-reported their speaking proficiency on a scale from 1 (poor) to 7 (high), as well as their degree of accentedness on a scale from 1 (weak) to 7 (strong). The column "Accent" refers to whether they thought they had more of an American or British accent.

| Subject | Time spent in an English-speaking country | English-speaking country in which you spent most time | Preferred varieties | Speaking proficiency | Degree of accentedness | Accent |
|------------|---|---|---------------------|----------------------|------------------------|----------|
| <i>S1</i> | 1 | USA | American | 6 | 3 | American |
| <i>S2</i> | 6 | England | British | 5 | 4 | American |
| <i>S3</i> | 18 | England | British | 5 | 2 | British |
| <i>S4</i> | 14 | Ireland | British | 5 | 4 | American |
| <i>S6</i> | 6 | USA | British | 6 | 5 | American |
| <i>S7</i> | 9 | England | British | 6 | 4 | British |
| <i>S8</i> | 7 | England | British | 5 | 3 | British |
| <i>S10</i> | 5 | Canada | American | 6 | 2 | American |
| <i>S13</i> | 13,5 | South Africa | British | 7 | 2 | British |
| <i>S14</i> | 7 | England | American | 4 | 5 | British |
| <i>S15</i> | 2 | USA | American | 6 | 5 | American |
| <i>S16</i> | 0,25 | England | American | 5 | 4 | American |
| <i>S17</i> | 0,75 | USA | British | 5 | 3 | British |
| <i>S20</i> | 2 | USA | British | 5 | 4 | American |
| <i>S22</i> | 2 | Canada | British | 7 | 4 | American |
| <i>S23</i> | 1,75 | USA | British | 7 | 3 | American |
| <i>S24</i> | 0,25 | England | British | 5 | 2 | British |
| <i>S27</i> | 10,5 | England | British | 7 | 2 | British |
| <i>S28</i> | 1,25 | England | British | 6 | 3 | American |
| <i>S29</i> | 0,5 | England | British | 5 | 4 | British |

U^S-PC

Université Sorbonne
Paris Cité

université
PARIS
PARIS 7
DIDEROT



Université de Paris

QUESTIONNAIRE INITIAL

Identifiant :

Adresse e-mail :

Année de naissance :

Lieu de naissance :

Langue(s) maternelle(s) :

Langue(s) maternelle(s) du père :

Langue(s) maternelle(s) de la mère :

Lieux de résidence (jusqu'à 10 ans). Veuillez préciser la durée si vous avez résidé dans plusieurs lieux différents :

Autre(s) lieu(x) de résidence (après 10 ans). Veuillez préciser la durée si vous avez résidé dans plusieurs lieux différents ;

Lieu de résidence actuel :

Fumez-vous ? Entourez la réponse qui correspond à votre choix :

OUI

NON

L'anglais est votre (entourez la réponse qui correspond à votre choix) :

deuxième langue troisième langue autre (précisez)

Combien de temps avez-vous passé au total dans des pays anglophones ? Veuillez indiquer votre réponse en mois soit : 0.25 = une semaine, 1 = 1 mois, 12 = un an, etc.

Dans quel pays anglophone avez-vous passé le plus de temps ? (une seule réponse possible)

Dans quelle ville anglophone avez-vous passé le plus de temps ? (une seule réponse possible)

Combien de temps y êtes-vous restée ?

Avez-vous, à votre connaissance, des antécédents de troubles de la voix ou de l'audition ? Si oui, veuillez préciser lesquels :

Combien de temps par semaine passez-vous au contact de la langue anglaise, et dans quelles circonstances majoritairement (en cours, avec des amis ou des membres de la famille, loisirs...)?

Combien de temps par semaine passez-vous à lire en anglais :

Combien de temps par semaine passez-vous à écrire en anglais :

Combien de temps par semaine passez-vous à parler anglais :

Combien de temps par semaine passez-vous à écouter des émissions anglophones ou regarder des programmes anglophones :

Quand vous écoutez de l'anglais, est-ce plutôt de l'anglais britannique ou américain ? Entourez la réponse qui correspond à votre choix :

Britannique

Américain

Préférez-vous les accents britanniques ou américains ? Entourez la réponse qui correspond à votre choix :

Britanniques

Américains

AUTO-EVALUATION

Comment jugez-vous votre maîtrise de la langue anglaise ? Entourez la réponse qui correspond à votre choix :

Mauvaise 1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 Très bonne

Considérez-vous votre accent comme étant plutôt britannique ou plutôt américain ? Entourez la réponse qui correspond à votre choix :

Britannique

Américain

Votre accent est-il (entourez la réponse qui correspond à votre choix) :

Très américain

Plutôt américain

Neutre

Très britannique

Plutôt britannique

Comment estimez-vous votre degré d'accent lorsque vous parlez anglais ? Entourez la réponse qui correspond à votre choix :

Léger 1 ___ 2 ___ 3 ___ 4 ___ 5 ___ 6 ___ 7 Fort

Appendix D

IRB Requirements

Présidente : Jacqueline Fagard

**N° 2020 - 08-
BALLIER-BURIN.**

PROTOCOLE : Acoustic and electroglottographic study of laryngealisation: phonological and paralinguistic adaptation among French learners of English

Noms du/des chercheur(s) : Léa burin ; Nicolas Ballier

Email pour la correspondance : lea.burin@univ-paris-diderot.fr; nicolas.ballier@univ-paris-diderot.fr

Labo/Service : Centre de Linguistique Inter-langues, de Lexicologie, de Linguistique Anglaise et de Corpus-Atelier de Recherche sur la Parole, Université Paris-Diderot

Évalué à la séance du 10/03/2020

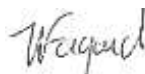
AVIS : Favorable

Par cet avis favorable, le CER U-Paris ne se prononce pas sur le respect des mesures barrières contre le Covid-19. Afin de protéger les participants et les personnes impliquées dans la recherche, les responsables de la recherche doivent impérativement se mettre en conformité avec les mesures préconisées pour toutes recherches sur site et hors site par les tutelles hébergeant les unités de recherche concernées, via le DU et responsable hygiène et sécurité.

L'avis favorable du CER U-Paris n'exonère pas des formalités réglementaires. A cet égard, il vous appartient notamment, si vous traitez des données se rapportant à un individu directement ou indirectement identifiable, de vous conformer au règlement européen sur la protection des données (RGPD) en vigueur depuis 2018. Pour cela, vous pouvez solliciter les conseils du Correspondant informatique et libertés (CIL) ou du service juridique de votre université ou de votre organisme de recherche.

Les investigateurs faisant appel aux services du CER U-Paris s'engagent à lui signaler tout événement non anticipé survenant en cours d'étude. Ces éléments seront utilisés aux fins d'amélioration des futurs services et conseils que le CER U-Paris pourrait donner.

N° IRB : 00012020-08



Jacqueline Fagard



Université Sorbonne
Paris Cité



Notice d'information

Étude linguistique - EGG

CLILLAC-ARP, 8 Place Paul Ricoeur
75013 Paris Diderot

Responsables: Nicolas Ballier, nicolas.ballier@u-paris.fr et Emmanuel Ferragne, emmanuel.ferragne@u-paris.fr
Investigatrice principale: Léa Burin, lea.burin@u-paris.fr

Madame,

Vous avez accepté d'envisager la participation à cette étude et nous vous en remercions. Celle-ci est destinée à étudier la manière dont les apprenantes francophones acquièrent la prononciation anglaise.

Cette recherche est sans bénéfice individuel direct. Cependant, elle permettra à terme de mieux comprendre la manière dont certaines caractéristiques linguistiques sont acquises. L'étude ne présente pas de risques connus.

L'étude comprend une simple séance lors de laquelle plusieurs tâches seront effectuées. Les tâches dites de production seront conduites avec un électroglottographe (EGG).

L'électroglottographie est une technique de mesure de la surface d'accolement des cordes vocales au cours de la phonation. Cette technique, non-invasive, consiste à placer un collier muni de deux électrodes qui seront placées de part et d'autre du larynx. Un gel permettant une meilleure conductivité aura été appliqué sur les électrodes en amont. Cette technique ne présente aucun danger particulier, ni contre-indication connue.

Vous serez assises, munie d'un micro-casque et appareillée avec un électroglottographe, dans une chambre insonorisée. La première tâche consistera en la lecture de phrases. Vous entendrez, par la suite, ces mêmes phrases produites par un locuteur natif que vous devrez imiter.

La tâche d'évaluation consistera en l'écoute de plusieurs stimuli. Ces derniers devront être évalués sur une échelle de 1 à 6 en fonction de différents traits qui vous seront communiqués.

La totalité ne devrait pas excéder une heure.

Les informations recueillies sont la propriété de l'Université de Paris et seront traitées confidentiellement. Elles pourront faire l'objet de présentations scientifiques ou de publications dans la presse scientifique, mais dans aucun cas votre nom ne figurera avec les données.

Votre participation à cette étude est volontaire et pourra être interrompue à tout moment sur simple demande de votre part, sans avoir à apporter de justification. Le refus de participer à cette étude n'aura aucune conséquence sur la bonne poursuite de vos études universitaires.

Pendant la durée de l'expérience, et après celle-ci, vous pourrez contacter l'expérimentatrice, Mme Léa Burin, pour toute information supplémentaire et pour avoir un retour sur les résultats de l'expérience.

Toutes les informations vous concernant seront conservées de façon anonyme et confidentielle. Les résultats pourront faire l'objet de communications dans des congrès et de publications scientifiques mais ne seront pas utilisés dans un autre cadre que celui décrit, et aucune information permettant de vous identifier ne sera révélée.

Vous recevrez une compensation de 15€ pour une participation complète aux séances d'enregistrement et d'évaluation.

Responsables étude : Emmanuel Ferragne & Nicolas Ballier

ETUDE SUR L'ACQUISITION DE LA PRONONCIATION ANGLAISE FORMULAIRE DE CONSENTEMENT

Consentement du participant à l'étude :

Je soussignée, certifie avoir été informée du déroulement de l'étude.

En particulier :

- j'ai lu la notice d'information et j'ai eu l'opportunité de discuter avec l'expérimentatrice et de lui poser toutes les questions que je me posais au sujet de l'expérience.
- je donne mon autorisation pour que ma voix soit enregistrée. Ces données seront stockées sur un disque dur prévu à cet effet. Seuls les responsables d'étude, ainsi que l'expérimentatrice auront accès à ces données.
- je sais que toutes les informations recueillies ici ne seront pas diffusées à une tierce personne, elles seront utilisées avec confidentialité, uniquement dans le cadre de cette étude et de manière anonyme.
- je suis prête à suivre les instructions données par l'expérimentateur, avant, pendant, et après l'expérience. Si je ne suis pas les instructions, l'expérimentateur pourra interrompre ma participation dans l'étude.
- j'ai le droit d'interrompre ma participation à l'expérience à tout moment et dans n'importe quelle phase de l'expérience.
- j'ai le droit de demander la suppression de mes données à tout moment.

Date et signature :

Email et numéro de téléphone :

Si vous avez des questions au sujet de l'étude, merci de contacter l'expérimentatrice principale Léa Burin (CLILLAC-ARP, lea.burin@u-paris.fr)

Résumé

Perception et accommodation chez les apprenants français de l'anglais :

Une étude acoustique et électroglottographique de la voix craquée

Nous avons analysé la faculté des apprenantes françaises à imiter la voix craquée produite par des locuteurs français ou américains. L'influence du genre a été analysée au sein d'une même langue. Nous avons adopté une double approche qui comprend à la fois des mesures acoustiques et électroglottographiques (EGG), ainsi qu'une évaluation perceptive de la qualité de la voix.

Notre étude comparative de la voix craquée française et américaine a démontré que toutes les variables ne rendaient pas compte du même effet selon la mesure analysée. Il y a un effet global de la langue sur CPP et f_0 , un effet du genre sur f_0 et $H2^*-H4^*$, et du genre en français sur PIC. Les deux mesures EGG n'ont montré que peu de variation dans les deux langues. Davantage de variabilité a été observée dans la voix craquée américaine. Les différences hommes/femmes sont plus importantes en anglais américain, avec une plus grande variabilité observée dans la façon dont les femmes produisent la voix craquée. Aucune différence notable entre les sexes n'a été observée en français. Il est possible qu'il existe plus de sous-types de voix craquée en anglais américain, les femmes en produisant davantage. Ces différences au sein du genre confirment que l'utilisation de la voix craquée est un phénomène social, et que ce dernier est plus répandu en anglais américain.

Un plus grand nombre d'occurrences de craquement en position finale a été observé lorsque les sujets lisaient en anglais. Les voyelles basses ont été plus fréquemment craquées que les hautes, dans les deux langues. Aucun effet de convergence globale n'a été observé. Seule la durée des voyelles converge globalement, tandis que $H1^*$, $H2^*$, $H1^*-A2^*$ et $H1^*-A3^*$ divergent globalement. Les dimensions recevant le plus d'effets dans les deux langues diffèrent. Aucun effet significatif du genre n'a été observé au sein d'une même langue.

Les deux qualités de voix ont été évaluées de manière similaire. La voix craquée a été évaluée légèrement plus négativement, surtout, chez les modèles français et inversement chez les locuteurs américains. Elle a aussi été jugée de manière plus positive si elle est produite par les locuteurs américains plutôt que par les locuteurs français, et, globalement, plus positivement quand elle est produite par les femmes. Elle a été jugée plus positivement que la voix non-craquée lorsqu'elle est produite par les hommes américains. Pratiquement aucune différence n'a été observée entre les deux qualités de voix dans les enregistrements des femmes américaines. Elle a été jugée plus positivement avec les voix de femmes françaises qu'avec les voix des hommes français. Elle a toujours été évaluée plus négativement que la voix non-craquée en français, qu'elle soit produite par les hommes ou par les femmes.

Mots-clés: voix craquée, accommodation, électroglottographie, perception/production de la qualité de voix, anglais L2

Abstract

Perception and accommodation among French learners of English: An acoustic and electroglottographic study of creaky voice

We analysed accommodation of creaky voice from female French speakers towards both French and American English speakers. Cross-gender differences were analysed within-language. We adopted a two-fold approach that included both acoustic and electroglottographic (EGG) measures of convergence, and a perceptual evaluation of voice quality.

Comparison of French and American English creaky voice showed that not all variables accounted for the same effect depending on the measurement that was being analysed. CPP and f_0 distinguished creaky vowels between languages, f_0 and H2*-H4* distinguished creaky vowels across gender, and cross-gender differences were only observed for PIC in French. Both EGG measures displayed very little variance in both French and American English. f_0 and H1*-H2* were found to vary more in American English creaky voice. Cross-gender differences were greater in American English, with more variability observed in the way female speakers produced creaky voice. No striking cross-gender differences were observed in French. More sub-types of creaky of creaky voice might exist in American English, with female producing more of them. These cross-gender differences also confirm that creaky voice usage is socially constructed and is more predominant in American English.

More creaky voice occurrences were observed on word-final position when subjects read in English. Low vowels received more creak than high vowels, in both languages. Convergence analysis showed no overall significant convergence effects. Only vowel duration received overall convergence effects, while H1*, H2*, H1*-A2* and H1*-A3* received overall divergence effects. We observed a language effect on some dimensions, with overall more convergence and less divergence towards/from American English speakers. There was considerable variability as to what dimensions received more accommodation effects across language. No significant gender effect was observed within-language.

Both voice qualities were similarly evaluated, with non-creaky voice slightly more positively. We observed more positive judgment for creaky vs. non-creaky voice when produced by American English speakers, while non-creaky voice was rated more positively when produced by French speakers. Creaky voice was more positively rated when produced by American English than by French speakers. Cross-gender analyses revealed more positive judgment towards female than towards male creaky voice. Creaky voice was more positively judged when produced by male American English. There was barely any difference in ratings of creaky vs. non-creaky voice when produced by female American English speakers. Creaky voice was more positively rated when produced by female than by male French speakers. However, creaky voice was always rated more negatively than non-creaky voice, whether produced by male or female French speakers.

Keywords: creaky voice, accommodation, electroglottography, voice quality perception/production, L2 English

Le larynx est une structure complexe faite d'os, de cartilages et de muscles. Certains ont un rôle particulièrement important dans le processus de phonation. Parmi eux, on retrouve par exemple les cartilages thyroïde, cricoïde et aryténoïdes (Jones & Barnes, 2019). Ces cartilages sont connectés les uns aux autres par différents muscles. Chaque muscle est nommé à partir des structures qu'il relie (e.g. inter-aryténoïdien, crico-aryténoïdien, etc.). Le niveau de tension émis sur ces différents muscles et cartilages va modifier la longueur, l'épaisseur, et le degré d'approximation des plis vocaux. Selon Laver (1980), les différentes configurations laryngées, et donc les différents types de phonation/qualités de voix, peuvent être définis et catégorisés selon trois paramètres : la tension adductive, la tension longitudinale, et la compression médiale⁷¹.

Pour être produite, la voix craquée nécessite une tension adductive et une compression médiale conséquentes, mais une tension longitudinale moindre (Wright *et al.*, 2019). Une faible pression sous glottique et un faible débit d'air caractérisent souvent cette qualité de voix (Podesva, 2013). Les plis vocaux vibrent très lentement et passent plus de temps accolés que disjoints (Johnson, 2013). Néanmoins, il n'existe pas un mais plusieurs types de voix craquée, chacun étant représenté par un ensemble de caractéristiques acoustiques bien précis (Keating *et al.*, 2015). Le type que nous venons de décrire correspond à la voix craquée dite prototypique. Il existe cinq autres sous-types, mais nous n'avons pas eu recours à un tel niveau de granularité dans notre étude puisque la tâche de classification s'est avérée particulièrement épineuse, chronophage et difficile à conduire pour certains cas ambigus.

Les types phonatoires n'ont pas nécessairement les mêmes usages dans les langues du monde. En anglais américain, la voix craquée peut être utilisée comme allophone de la voix modale sur les voyelles précédant une consonne occlusive non-voisée (Gordon & Ladefoged, 2001). Elle est également utilisée pour marquer les frontières prosodiques, plus particulièrement les débuts et fins de phrase (Umeda, 1978; Kreiman, 1982; Pierrehumbert & Talkin, 1992; Pierrehumbert, 1995; Dilley *et al.*, 1996; Redi & Shattuck-Hufnagel, 2001; Epstein, 2003; Ni Chasaide & Gobl, 2004; Slifka, 2006; Wolk *et al.*, 2012; Podesva, 2013; Garellek, 2014), et survient fréquemment sur les syllabes accentuées (Pierrehumbert, 1995; Dilley *et al.*, 1996). Les types phonatoires

⁷¹Les différentes configurations laryngées sont, en réalité, réalisées de manière bien plus complexe. Elles impliquent notamment les structures musculaires de la langue et du pharynx (Catford, 1964; Esling *et al.*, 2019). Nous les avons délibérément exclus de notre étude, cette dernière s'intéressant essentiellement aux événements glottiques.

peuvent aussi avoir des fonctions paralinguistiques (humeur et attitude) et sociolinguistiques. En anglais américain, la voix craquée peut être employée pour paraître plus autoritaire (Dilley *et al.*, 1996; Lefkowitz & Sicoli, 2007) ou bien à des fins stylistiques, notamment dans le but de charmer (Pennock, 2015) ou de renvoyer l'image d'une femme indépendante (Yuasa, 2010). Il n'y a aucune mention de la voix craquée en français en dehors de marqueurs d'hésitation (Honikman, 1964).

Cette étude porte partiellement sur l'accommodation de la voix craquée. La convergence phonétique est le processus qui désigne un rapprochement des caractéristiques vocales de locuteurs en interaction. D'un point de vue psycholinguistique, la convergence est considérée comme automatique et cognitive. C'est une conséquence inévitable de la manière dont la langue est traitée dans le cerveau et résulte d'un lien direct entre la perception et la production (Sancier & Fowler, 1997 ; Goldinger, 1998). Les études portant sur l'accommodation interlangue rendent souvent compte d'un effet du statut de la langue (Zajac, 2013), de la manière dont les sons présents dans la L2 sont perçus (Best & Tyler, 2007), et du niveau de langue (Wang & Gu, 2022) ou du 'talent phonétique' (Lewandowski & Jilka, 2019). D'un point de vue sociolinguistique, la convergence est étroitement liée à certaines considérations sociales, telles que le genre du locuteur et de l'interlocuteur (e.g. Babel, 2012), de la dominance interactionnelle (e.g. Andreeva *et al.*, 2021), de l'affiliation culturelle (e.g. Babel, 2010) ou encore de la préférence sociale (e.g. Babel, 2009). D'un point de vue hybride, certains facteurs perceptivo-moteurs et certaines influences sociales et/ou psychologiques sont conjointement actifs lorsqu'une convergence est observée.

Ici, nous nous intéressons à l'effet de la langue et du genre (homme/femme) sur la perception et sur l'accommodation de la voix craquée chez des locutrices françaises apprenantes de l'anglais. Nos hypothèses, qui s'appuient sur des études antérieures, sont les suivantes :

1. Pour l'accommodation, nous nous attendons à une convergence plus importante vers les locuteurs américains, ainsi que vers les locuteurs masculins.
2. Pour la perception, nous nous attendons à une évaluation plus positive de la voix craquée lorsque celle-ci est produite par les locuteurs américains. L'effet du genre sur l'évaluation de la voix craquée reste à déterminer par l'expérience.

Nous avons constitué un corpus contrôlé inédit afin de mesurer la voix craquée et non-craquée dans des situations comparables. Les stimuli reposent sur des phrases courtes de 5 à 7 syllabes enregistrées par 3 hommes américains, 3 femmes américaines, 3 hommes français, et 3 femmes françaises. Seuls les exemples de voyelles en position finale de l'unité prosodique ont été prises en compte dans l'analyse (voyelle contenue dans la dernière monosyllabe accentuée). Ces dernières étaient produites à la fois avec une occurrence de craquement, ou sans (soit une même phrase mais deux qualités de voix différentes). Nous avons uniquement utilisé des phrases déclaratives puisque ces dernières induisent un ton bas et descendant dans les deux langues. Les sujets sont 20 locutrices cisgenres ayant le français pour langue maternelle et qui étudient l'anglais à Université Paris Cité. Elles ont pris part à trois tâches différentes : une tâche de lecture dans laquelle elles ont lu les mêmes phrases que celles enregistrées par les locuteurs modèles, une tâche de répétition/imitation dans laquelle on leur a demandé de répéter et d'imiter du mieux qu'elles pouvaient la production des locuteurs modèles qui leur était présentée, et une tâche d'évaluation/jugement dans laquelle elles ont dû évaluer la voix du locuteur modèle qui leur était présentée selon 4 caractéristiques différentes ("agréable, attrayante, puissante et éduquée"). A chaque caractéristique correspondait une échelle de Likert comprenant 6 choix de réponse allant de 'très' à 'pas du tout'. Nous avons utilisé les mêmes adjectifs que ceux utilisés par Bayard *et al.* (2001), et plus particulièrement ceux qu'ils définissent comme "caractéristiques de la qualité de voix".

Les tâches dites de production ont été conduites avec un électroglottographe. L'électroglottographie est une technique de la mesure de la surface d'accolement des plis vocaux au cours de la phonation. Cette technique est non-invasive et consiste à appareiller le sujet d'un collier muni de deux électrodes qui sont positionnées de part et d'autre du larynx, au niveau du cartilage thyroïdien. Un gel permettant une meilleure conductivité aura été appliqué sur les électrodes en amont. Cette technique ne présente aucun danger particulier, ni contre-indication connue. Nous avons mesuré acoustiquement la durée, la fréquence fondamentale, l'amplitude des deux premiers harmoniques et leur différence corrigée, quatre autres différences d'amplitude harmonique, et le pic de prééminence spectral (CPP). A ces mesures s'ajoutent deux quantifications relatives aux signaux électroglottographiques : le quotient fermé et le PIC (pic d'augmentation du contact des

plis vocaux).

Nous avons premièrement analysé et comparé les mesures citées en fonction de la langue et du genre dans le but d’observer (si tel est le cas) quelles dimensions permettent de distinguer la voix craquée française de la voix craquée américaine. Les principaux résultats montrent que toutes les dimensions ne rendent pas compte du même effet selon la mesure analysée. La voix craquée française semble être produite avec une fréquence fondamentale significativement plus haute que l’américaine et avec des valeurs de CPP significativement plus basses. Les différences observées entre les deux langues semblent indiquer que la voix craquée française serait produite avec une glotte moins fermée que la voix craquée américaine. Quand elle est produite par les hommes, la voix craquée prend des valeurs de f_0 plus basses et des valeurs de H2*-H4* plus hautes. Nous n’avons observé un effet significatif du genre au sein d’une même langue qu’en français : les femmes produisent les voyelles craquées avec des valeurs de PIC plus basses. Plus les valeurs de PIC sont élevées, plus cela témoigne d’une voix dite “soufflée” (Esposito, 2012). Il semblerait donc que la voix craquée soit produite avec une glotte moins fermée lorsqu’elle est produite par les hommes français (que par les femmes françaises).

Nous avons conduit une analyse en composantes principales (ACP) dans le but d’observer le rôle éventuel du genre dans la variance observée au sein d’une même langue. Nous avons observé une variabilité relativement importante dans la manière dont la voix craquée a été produite par les femmes américaines. Les différences entre les genres observées en français sont, quant à elles, insignifiantes. Nous supposons que certaines différences sont dépendantes de la langue et que la physiologie seule ne peut expliquer les différences acoustiques observées dans la production de la voix craquée. Il est possible qu’il existe plus de sous-types de voix craquée en anglais américain, les femmes produisant davantage de types de voix craquée différents. Ces observations confirment que l’utilisation de la voix craquée est un phénomène social, et que ce dernier est plus répandu en anglais américain.

Nous avons analysé et comparé l’usage de la voix craquée lorsque les sujets lisaient en français (L1) et en anglais (L2). Les apprenantes produisent plus d’occurrences de craquement lorsqu’elles lisent en anglais, ce qui corrobore le fait que l’usage de la voix craquée est dépendant du statut de la langue (Pillot-Loiseau *et al.*, 2019). Notre corpus témoigne d’une forte variabilité inter-

individuelle, mais nous n'avons observé aucune corrélation entre la proportion de craquement en français (L1) et en anglais (L2). Malgré le fait que le nombre d'occurrences diffère grandement par catégorie de voyelle, il semblerait que les voyelles basses soient plus fréquemment craquées que les voyelles hautes, et ce, dans les deux langues, ce qui corrobore les observations de Panfili (2015).

Pour évaluer les effets de convergence, nous avons utilisé une mesure souvent employée dans les études portant sur l'accommodation : la *difference-in-distance* (Babel, 2012; Pardo *et al.*, 2013, 2017; Lewandowski & Nygaard, 2018, Wagner *et al.*, 2021). Cette mesure de distance a été calculée pour chaque occurrence et pour chaque mesure acoustique et EGG. Les principaux résultats montrent une variation importante dans le processus d'accommodation, mais aucun effet de convergence globale n'a été mis en évidence, à l'exception de la durée des voyelles. Toutefois, la convergence d'ensemble vers les locuteurs américains semble plus nette. Les dimensions recevant une convergence ou une divergence varient en fonction de la langue. Nous avons aussi observé une forte variabilité entre les genres au sein d'une même langue. Pour l'anglais américain, il y a eu un effet de convergence sur le même nombre de dimensions lorsque les sujets répétaient les hommes et les femmes, mais aussi un effet de divergence sur un nombre plus important de dimensions lorsque les sujets répétaient les hommes. Pour le français, nous avons observé une convergence sur un plus grand nombre de dimensions lorsque les sujets répétaient les hommes et une divergence sur un plus grand nombre de dimensions lorsque les sujets répétaient les femmes. Un effet du locuteur modèle a été observé pour la plupart des dimensions.

Nous avons comparé l'évaluation de la voix craquée à celle de la voix non-craquée en fonction de la langue et du genre au sein d'une même langue. Les principaux résultats de cette analyse perceptive montrent que la voix non-craquée a été dans l'ensemble évaluée plus positivement. La voix non-craquée semble être mieux perçue que la voix craquée en français, alors que le contraire s'observe en anglais. En anglais américain, les deux qualités de voix ont été évaluées de manière similaire quand elles sont produites par les femmes, mais la voix craquée a été évaluée de manière plus positive que la voix non-craquée lorsque produite par les hommes. De manière globale, en anglais américain, la voix craquée a aussi été plus positivement évaluée quand elle est produite par les femmes que quand elle est produite par les hommes. En français, la voix non-craquée a

été plus positivement évaluée que la voix craquée, qu'elle soit produite par les hommes ou par les femmes. La voix craquée a aussi été évaluée de manière plus positive quand elle est produite par les femmes, dont les voix ont, globalement, été plus positivement évaluées que celles des hommes.

Nous avons comparé l'évaluation de la voix craquée à celle de la voix non-craquée pour chaque caractéristique de manière indépendante. Une fois de plus, nous avons considéré l'influence de la langue et du genre. De manière globale, la voix non-craquée a été plus positivement évaluée que la voix craquée. La voix craquée a seulement été évaluée comme étant plus attrayante que la voix non-craquée. En anglais américain, les caractéristiques *attrayante* et *éduquée* ont reçu une évaluation plus positive lorsque les locuteurs utilisaient la voix craquée. La voix non-craquée a été évaluée comme plus puissante que la voix craquée, et la voix craquée semble être perçue comme étant aussi agréable que la voix non-craquée. En français, toutes les caractéristiques ont reçu une évaluation plus positive lorsque les locuteurs utilisaient la voix non-craquée, avec des résultats significatifs pour *puissante* et *éduquée*. L'influence du genre en anglais américain a démontré une évaluation plus positive pour les caractéristiques *agréable*, *attrayante*, et *éduquée* lorsque les locuteurs hommes utilisaient la voix craquée. Les caractéristiques *attrayante* et *éduquée* ont reçu une évaluation plus positive lorsque les locutrices modèles utilisaient la voix craquée, mais les deux autres caractéristiques – *puissante* and *éduquée* – ont reçu une évaluation plus positive lorsque les locutrices modèles utilisaient la voix non-craquée. En français, toutes les caractéristiques ont reçu une évaluation plus positive lorsque les locuteurs, hommes et femmes confondus, utilisaient la voix non-craquée. Le fait que nous ayons observé un effet quasi systématique du locuteur modèle dans la tâche de répétition/imitation nous a conduit à également prendre en compte cet effet dans notre analyse perceptive. Nous avons observé peu de variation en fonction des différents modèles et de la qualité de voix utilisée. Un modèle en particulier, FF_08, a reçu une évaluation significativement plus positive, indépendamment de la qualité de voix utilisée, et ce pour chaque caractéristique. Nous pouvons conclure que, quand ils jugent une voix, les locuteurs ne prennent pas seulement en compte la qualité de voix mais probablement un nombre important de caractéristiques de la voix et du discours.