The other-accent effect on speaker recognition: A study on Quebec and Hexagonal French[1]

Pamela Bautista-Boivin[2]
Université du Québec à Montréal, Montréal, Canada

Julien Plante-Hébert
Université du Québec à Montréal, Montréal, Canada

Abstract: The present article investigates the other-accent effect (OAE) on speaker recognition in the context of voice line-ups for speakers of Quebecois and Hexagonal (France) French. The literature largely attests to a language familiarity effect (LFE) that can bias the results of this forensic phonetics technique. A far less substantial number of studies have investigated whether this finding also extends to varieties of a single language (regional or social). The main aims of the present study are therefore to test whether such an effect is present for the two varieties of French concerned, and whether the predominance of the so-called "standard" variant of French generates a measurable asymmetry in this effect. Participants (n = 34) whose native French was either Quebecois or Hexagonal took part in a speaker recognition task through two voice line-ups, one for each variety of French. The findings indicate that there is no significant OAE on speaker recognition for the French varieties studied, despite some noteworthy tendencies related to the asymmetry between the two varieties of French and the duration of stay of the French participants in Quebec.

Keywords: speaker recognition; voice line-up; other-accent effect; forensic phonetics

1 Introduction

The ability to recognize faces or voices is a fascinating human social skill that has been the subject of numerous sociological and psychological studies (Doty, 1998). We perform this type of cognitive task on a daily basis without really being aware of it (Goldstein et al., 1981). Nevertheless, the accuracy and fallibility of facial and voice recognition are not yet fully defined (Atkinson, 2015; Harvey et al., 2023). Following the work of Stevenage and Neil (2014), it is possible to assume that the recognition of faces and voices share a number of similarities and interact on multiple levels. The authors, however, stress that visual identity processing is much more robust and efficient than its auditory counterpart. Visual identity processing has also received much more scientific attention. Among the research that has focused on these topics, a few have

[1] This article is a translation of Bautista Boivin, P., & Plante-Hébert, J. (2024). L'effet de l'accent étranger sur la reconnaissance de locuteur.trice.s : étude du français québécois et hexagonal. *Criminologie, Forensique Et Sécurité*, *2*(2). https://doi.org/10.26034/la.cfs.2024.4688

[2] *Corresponding Author: bautista_boivin.pamela@courrier.uqam.ca*

 ISSN 2564-2855

nonetheless managed to establish that familiarity with a person tends to greatly improve performance on a voice identification task, to a level that would surpass even the capabilities of automated systems (Atkinson, 2015; Plante-Hébert & Boucher, 2014; Yu et al., 2021). But what about voices with which we are not familiar? In addition to providing answers that contribute to a better understanding of the processes underlying speech processing, this is the type of question that is of particular interest to experts in forensic phonetics.

Forensic phonetics is a subfield of forensic linguistics, a line of scientific research that applies linguistic analyses and language methods to legal and investigative contexts (May et al., 2020). When these analyses involve acoustic material, either as evidence or for investigative purposes, they are considered to fall within the remit of forensic phonetics. These include, among others, the analysis of voice recordings, decoding of recorded messages, analysis of emotions in the voice, and the authentication of recordings (Hollien, 2012; Jessen, 2008, 2020; Rose, 2002). Over the last few decades, an exponential number of human interactions have been recorded by the means of new technologies that are constantly emerging. There has also been an increase in the number of language crimes perpetrated or captured by mobile devices, namely phone threats, ransom demands, frauds, or incriminating recordings of all kinds (Nolan, 1991; Statistics Canada, 2022; Watt & Brown, 2020).

As with any type of crime, the more information law enforcement agencies have on a crime suspect, the easier their investigation becomes. The forensic phonetics expert can provide law enforcement professionals with the benefits of their expertise by carrying out one of three types of analysis: speaker profiling, speaker recognition, or speaker identification. Analysis of the phonetic characteristics of the voice in question has one of three aims:

- in the absence of a suspect, profile the speaker according to characteristics such as age, biological sex, place of birth or current residence, level of education, etc. (speaker profiling; Jessen, 2020; Kreiman & Sidtis, 2011; Rose, 2002);
- compare a questioned voice sample with samples for which the identity is known in an attempt to confirm the identity of the suspect (speaker recognition; Watt & Brown, 2020);
- determine the exact identity of a person from an available voice sample (speaker identification).

These types of analyses can be carried out directly by the phonetics experts, using sophisticated techniques involving prosodic and articulatory measures such as fundamental frequency (F0), formants, articulation rate, and speech rhythm. In this article, we will focus on recognition by so-called naive individuals (Rose, 2002). Suppose an ongoing investigation concerns a bomb threat call: would it be possible for the individual who answered this call to recognize the caller's voice, throughout the legal procedures, for example, even though they only heard it once? Several variables can influence the accuracy of unfamiliar voice recognition, and to date, its reliability has yet to be proven (Atkinson, 2015; Bull & Clifford, 1999; Rose, 2002). In order for a layperson to carry out a speaker recognition task, however, the contribution of experts in forensic phonetics remains necessary. At present, one of the most widely recognized protocols for conducting and studying speaker recognition is called "voice line-up." This technique is analogous to the visual identification (or eyewitness) line-up, more regularly used by police investigators. The latter involves presenting a witness with several side-by-side individuals with relatively similar physical attributes (or a series of photos of them) to determine whether it is

possible for them to recognize, among these individuals, the one they observed committing a crime (Public Prosecution Service of Canada, 2018). In a similar way, during a voice line-up, several acoustically similar voices are presented to an individual so that they can attempt to recognize or identify the voice of a suspect among them (Atkinson, 2015; Eriksson, 2007; Hollien, 2012; Jessen, 2008; Nolan, 2003; Philippon et al., 2007; Plante-Hébert & Boucher, 2014; Wilding et al., 2000).

1.1 Elaboration of a voice line-up

Although the use of voice line-ups, like forensic phonetics, is relatively new, several criteria have already been established to ensure that they are practised in a fair and legally valid manner (de Jong-Lendle et al., 2015; Harvey et al., 2023; McDougall, 2013). First, the speech samples must come from individuals who are similar in terms of age, gender, dialect or variety of language spoken, and pitch. They must also be controlled so that the prosody and length of the speech delivered are similar (Atkinson, 2015; Foulkes & French, 2012; Jessen, 2008; McDougall, 2013; Nolan, 2003; Plante-Hébert & Boucher, 2014; Stevenage et al., 2012). Secondly, the "target voice," i.e., the voice of the suspect, must not be too obviously distinct from the other voices included in the line-up, also known as "foil voices"; otherwise, the task is too easy. The opposite scenario can just as easily affect the validity of the demonstration, i.e., if the voices are so similar that it becomes impossible to distinguish the target voice (Atkinson, 2015; de Jong-Lendle et al., 2015; McDougall, 2013). Our review of the literature shows that a six-voice line-up is more commonly used and generally recommended (Cook & Wilding, 1997; Goggin et al., 1991; Harvey et al., 2023; Kerstholt et al., 2006; Thompson, 1987). The paper by Harvey et al. (2023) suggests that among voice line-ups composed of one, six, or ten voices, those composed of six voices represent a reasonable middle ground for optimizing successful recognitions. When the number of foil voices is too high, the number of false positives tends to increase (Harvey et al., 2023) and, of course, too few voices would lead to greater odds of successful recognition by chance.

1.2 Expertise effects

However meticulously elaborated, a line-up (whether vocal or visual) can still have blind spots. Various studies have demonstrated the existence of recognition biases in the context of visual line-ups, in particular a phenomenon known as the ethnicity effect (Atkinson, 2015; Hayward et al., 2008; Stevenage et al., 2012). The latter represents a subcategory of the expertise effect, which is defined as having a greater ease in processing a specific type of information due to greater exposure to it or more in-depth experience of it (Rhodes et al., 2006). The expertise effect is not unlike machine learning, whereby as a computer model is fed with more task-specific data, such as French linguistic data, for example, it improves and perfects its automatic language processing specifically for French (Bender et al., 2021). Indeed, the greater the access the model has to information reflecting French, the better it will be able to distinguish the word sequences that are most statistically probable or to make accurate predictions for the purposes of automatic speech recognition (ASR), machine translation, or text classification tasks in this language (Bender et al., 2021; Gastaldi, 2021). This linguistic model, however effective it may become in French, will nevertheless remain flawed when it comes to the world's other languages, since it will only have been trained on data specific to French.

The ethnicity effect is based on this principle. Although as human beings we have great expertise in distinguishing the physical features of thousands of individuals, there is a perceptual deficit regarding faces and voices from a different ethnic group from our own, with which we have lesser experience (Kuhl et al., 1992; Meissner & Brigham, 2001; Perrachione et al., 2010; Stevenage et al., 2012). This deficit translates into impaired recognition memory, accuracy of

gender and age judgment, and visual discrimination for faces of individuals from different ethnic groups (Atkinson, 2015; Hayward et al., 2008; Meissner & Brigham, 2001; Rhodes et al., 2006; Stevenage et al., 2012). For example, Caucasian individuals are better able to distinguish a Caucasian face among other Caucasian faces than an Asian face among other Asian faces (Hayward et al., 2008; Meissner & Brigham, 2001). Goldstein et al. (1981) also argue that the faces of people with whom we are not familiar become familiar more quickly if they belong to individuals from the same ethnic group as our own.

If the ethnicity effect is therefore a widely attested phenomenon at the visual level, the same cannot be said of its auditory counterpart. However, it turns out that a few researchers have actually observed a similar phenomenon: the language familiarity effect (LFE; Betancourt & Bahr, 2010; Goggin et al., 1991; Perrachione et al., 2010; Perrachione, 2018; Zarate et al., 2015). Essentially, it is easier for an individual to accurately identify voices and their characteristics in their native language than in a second or foreign language (Perrachione, 2018). Although several studies have investigated the impact of the LFE on speaker recognition, far fewer have delved deeper into this notion to ascertain whether this finding also extends to variants of the same language, be it on a regional or social scale. Nevertheless, most of those who did were able to demonstrate that an accent effect indeed occurred and that the difficulty for participants arose when tasked with recognizing voices that had not been produced in their own language variety compared to voices that belonged to the same variety as their own (Doty, 1998; Goggin et al., 1991; Kerstholt et al., 2006; Stevenage et al., 2012; Vanags et al., 2005; Yu et al., 2021).

Moreover, since the other-accent effect (OAE) is mostly based on the speaker's exposure to certain accents, it is not surprising that some researchers have observed an asymmetry in the way speakers are affected by it (Atkinson, 2015; Kerstholt et al., 2006; Perrachione et al., 2010; Stevenage et al., 2012; Vanags et al., 2005). It is common for a particular variety of a given language to be deemed more common (considered standard) because it is more widespread in the media, radio, television, and films, for example (Stevenage et al., 2012). Thus, speakers of less common varieties have more opportunities to perfect their perception of the "standard" accent since they are more exposed to it. Conversely, speakers of a "standard" variety of a language would have fewer opportunities to specialize in the fine processing of less widespread varieties. So, if we look at the effect of a different accent between a widespread and a more marginally spoken variety of the same language, it is expected that it will tend to be more pronounced for people whose mother tongue is the more widespread variety, while it is generally less pronounced for people whose mother tongue is a more marginal variety.

Finally, although a number of studies have investigated the OAE, most of them have focused on English (and sometimes Dutch or Spanish as well; Kerstholt et al., 2006; Thompson, 1987). As we develop knowledge of forensic phonetics in French-speaking Quebec, we feel it is necessary to determine the extent to which such accent effects are observed in French. Considering the significant French immigration to Quebec (Robitaille, 2023) and the marked distinction between the hexagonal (FH; spoken in France) and Quebecois (FQ) varieties of spoken French, the main objective of the present study is to replicate the observations made about the OAE using these two varieties of French. We expect this effect to be present in both FQ and FH. Assuming that this first hypothesis is verified, the second objective is to observe whether the asymmetry of this effect that some researchers have noted in their studies is observed for French speakers (Kerstholt et al., 2006; Perrachione et al., 2010; Stevenage et al., 2012). FH is widespread enough in Quebec that Quebec speakers have been widely exposed to it and have become accustomed to

it to some degree. Indeed, FH is often considered the "standard" French for the purposes of dubbing films and TV shows or second-language teaching (Kircher, 2012; Reinke & Ostiguy, 2012). There is also an important French population in Montreal that can contribute to the wider exposure of this variety of French (Consulat général de France à Québec, 2020). On the other hand, as the Quebec accent is less widespread in France, it would be reasonable to hypothesize that French speakers are less familiar with it. This is in line with the concept of the expertise effect mentioned above. In this sense, we postulate that native FQ participants will be less affected by the OAE than native FH participants. The third objective is to study the relationship between participants' level of certainty and their success on the recognition task. To this end, various studies have reported that confidence levels differ according to the accent of the voice participants are trying to identify (Atkinson, 2015; Goggin et al., 1991; Stevenage et al., 2012; Yu et al., 2021). We postulate that participants will be more certain of the validity of their response when it comes to recognizing their peers of the similar accent, regardless of whether they have managed to correctly identify the target voice or not. Our final research question examines the effect of the duration of stay in Quebec for participants of French origin on the OAE. We anticipate that those who have benefited from a longer exposure to FQ will show an attenuated OAE compared to those who have settled in Quebec more recently.

2 Method

2.1 Participants

Thirty-four participants aged between 27 and 45 were divided into two groups representing each of the accents studied: 18 people (10 women, 8 men) whose native language was FQ and who lived in Montreal, and 16 people (7 women, 7 men and 2 non-binary people) whose native variety of French was FH. The average age of the French was 33 years old (SD 6.27), while that of the Quebecers was 35 (SD 5.36). This age range was selected because, in the context of voice line-ups, recognition ability tends to decline with aging (Clifford et al., 1981; Eriksson, 2007). French participants were grouped into two categories according to the duration of their stay in Quebec: less than 5 years (n = 7) and 5 years or more (n = 9). None of the Quebec participants lived in France for more than a year. All participants confirmed that they had no known or diagnosed hearing problems. Lastly, participants had never heard the voices used as stimuli, to avoid the possibility that familiarity might facilitate recognition (Plante-Hébert et al., 2021). All participants read and signed an information and consent form approved by UQAM's *Comité institutionnel d'éthique de la recherche avec les êtres humains* (CIEREH).

2.2 Stimuli

Two voice line-ups (one in FQ and one in FH) were elaborated, each containing 1 target voice and 5 foils. Voice samples were recorded using the same procedure for each line-up. In order to control for the marked F0 variation that generally separates the sexes, only female voices were used as stimuli. At the time of recruitment, the following criteria were prioritized for the selection of the speakers: they should not have been diagnosed with, or have, any discernible language disorders such as stuttering or lisping (Atkinson, 2015; Jessen, 2008), and they should be speakers of either FQ or FH, with no marked regional accent. Given the difficulty in finding FH speakers living in Montreal that all came from only one specific region in France, FH speakers came from Bordeaux, Paris, Lille, Corrèze, Var and Luxembourg. The speaker from Var, though from the south of France, lived in Paris for 10 years and didn't have the typical southern accent. The speaker from Luxembourg, for her part, was educated in the French system. A French graduate student in

phonetics confirmed that none of the speakers' accents stood out. Regarding the FQ speakers, they mostly came from the Montreal region except two who came from Laurentides and Montérégie.

We did not try and elicit distinctions between the two varieties of French in the making of the stimuli, but rather expected that the most salient distinctions would appear naturally. Some phonetic markers are well known to distinguish between FH and FQ. Brasseur and Ménard (2013) showed that the laxing of high vowels /i/ and /u/ is one of the most perceptible markers followed by the affrication of alveolar plosives /t/ and /d/, the fronting of the nasal vowel /ɑ̃/ and the backing of the vowel /a/.

In a noise-proof room, on a desktop computer, and using a lapel microphone (Lavalier) connected to a 24-bit sound card (Scarlett 2i2 2nd Generation, Focusrite), each speaker was asked to record two texts as neutrally as possible, in terms of rhythm, intonation, and rate (Jessen, 2008). This was to minimize the effect of prosody on speaker identification (Stevenage et al., 2012). Recordings and signal processing were carried out using Audacity software (v3.2.1) at a sampling rate of 44.1 kilohertz and were saved in .wav format. Read speech was privileged in order to control for factors external to the voice, such as the use of discourse markers that might reveal the age, socioeconomic background, or educational level of the speakers, long pauses, hesitations, or any idiosyncrasies that could have caused noticeable variability within samples and thus undermine the validity of the voice line-up (Foulkes & Barron, 2000; Jessen, 2008; Stevenage et al., 2012). The two selected texts were devoid of excessively complex or specialized vocabulary and specific regional or dialectal expressions that could potentially not have been understood by participants from either region (Jessen, 2008). Poor understanding of the vocabulary used is a considerable factor influencing recognition (Goggin et al., 1991; Stevenage et al., 2012; Yu et al., 2021).

The recording of the first text (Appendix A) was intended for use in an exposure phase and had an estimated reading duration of 35 seconds. This duration was sufficient to ensure a reasonably high exposure time that would allow participants to access a more varied phonemic inventory during this phase (Atkinson, 2015; Cook & Wilding, 1997; Goggin et al., 1991; Harvey et al., 2023; Philippon et al., 2007). Indeed, the lack of linguistic information that results from listening to a recording that is too short can considerably impair a participant's ability to recall the target voice (Goldstein et al., 1981; Plante-Hébert & Boucher, 2015). Although all speakers produced recordings of this text, only the recordings of the target voices, determined a posteriori, were to be used during the experiment. The durations of the recordings of this text were 31.9 and 33.8 seconds respectively for the FH and FQ target voices selected.

The second text (Appendix B) was intended for the test phase, i.e., for the voice recordings included in the voice line-ups themselves. These lasted an average of 14.25 seconds each (SD 1.42 seconds). The literature on voice line-ups seems to agree that this duration is sufficient for speaker recognition (Atkinson, 2015; Broeders & van Amelsvoort, 1999; Goldstein et al., 1981; Harvey et al., 2023; Plante-Hébert & Boucher, 2015; Stevenage et al., 2012; Yu et al., 2021).

Each text was read aloud without recording a minimum of three times as practice, to enable the speakers to familiarize themselves with the texts and to make subsequent readings more natural and fluent. Three other readings were then recorded, from which only the best was retained.

2.2.1 Target voices

From the 8 speakers in each group of speakers (FQ and FH), the voice that appeared most neutral in terms of distinctiveness (speech rate, standard F0, neutral reading) was selected to be the target voice. This evaluation was carried out by ear by the authors of this article and by a

phonetically trained judge whose native variety of French is FH. For each of the two voice line-ups, five other voices were judged to be sufficiently neutral and similar to the target voice to be included among the foil voices, and two were disqualified on the grounds of distinctive resonances, overly pronounced accent, or hoarse quality.

2.3 Experimental task

The experimental task lasted around 20 minutes per participant, and took place in the presence of the experimenters in a quiet room. Each of the 34 participants was asked to perform two speaker recognition tasks, one for each variety of French studied. The speech samples were presented using circumaural headphones (HD 660S, Sennheiser) provided by the experimenters. Instructions and stimuli were presented using PsychoPy software (v2022.2.5) (Peirce et al., 2019). Participants began with the exposure phase, during which they were presented with a single recording of the target voice. They were instructed to focus their attention on the voice itself, rather than on the content of the text. Once they had finished listening, they were instructed that the test phase was about to begin and given written instructions. During this phase, the voice line-up composed of the target voice heard during the exposure phase and the 5 foil voices was presented. The order in which the voices were presented was fully randomized across participants. Each voice was associated with a number, which was displayed on the screen as it was presented. Participants were encouraged to take notes while listening, to help them make their selection later. Once all the voices of the same line-up had been heard, participants were instructed to indicate on a keyboard the number associated with the voice in the line-up corresponding to the one heard in the exposure phase to the best of their ability. A response was mandatory. The final step was to report their level of confidence in the answer provided on a Likert scale of 1, not confident, to 4, very confident, which was displayed on the screen. Once the responses for the first line-up were recorded, the process was repeated for the second. The order in which the line-ups were presented was balanced across participants, so that half heard the Quebec line-up first and vice versa.

2.4 Analysis

Data from 68 trials, i.e., 2 trials per participant, was collected. Out of these, 16 were removed as the target voice appeared first in the voice line-ups and all the target voices in this position were successfully recognized. Of the remaining 52 trials, 25 were from participants of French origin and the remaining 27 from Quebecers. In terms of the variety of French heard, 25 of the 52 trials selected corresponded to the FH voice line-up and the remaining 27 to that of the FQ.

Analyses of the results focused on success in the speaker recognition task, as well as on the confidence levels self-assessed by the participants after each voice line-up. The data collected on participant and speaker origins were used to determine accent similarity using two categories: similar accent and different accent. For example, the value “similar accent” was assigned to all trials in which a participant from Quebec was listening to the voice line-up featuring Quebec voices.

As success in the recognition task results in a binomial variable (success or failure), generalized mixed-effects models were used to analyze the data. With regard to the level of self-assessed confidence, mixed models were used. For all models, the only random factor was the participants’ ID and all models converged. All statistical analyses were performed using *jamovi* open-source software (*version 2.3.21*) (Fox & Weisberg, 2020; Gallucci, 2019; The jamovi project, 2023; R Core Team, 2021).

3 Results

3.1 The other-accent effect (OAE)

The first aim of the present study was to replicate the OAE, already attested in other languages, for FH and FQ. The average success rate in the recognition task was 69% (SE 9.2%) in the similar accent condition, and 65% (SE 9.5%) in the different accent condition (Figure 1, left). Using success on the recognition task as the dependent variable, we ran a generalized mixed model with the variable "similar accent" as the only fixed factor. The model was not significant meaning that the overall sample was not affected by an OAE. This result could be partly explained by the fact that participants from France spent a variable duration of time in Quebec.

Participants speaking both varieties of French showed a very similar level of confidence for the "similar accent" and "different accent" conditions, averaging 2.92/4 (SE 0.15) and 3/4 (SE 0.13). A mixed model with "similar accent" as the only fixed factor was also not significant meaning that the confidence of the participants was not statistically affected by the similarity of the accents of the speakers heard.

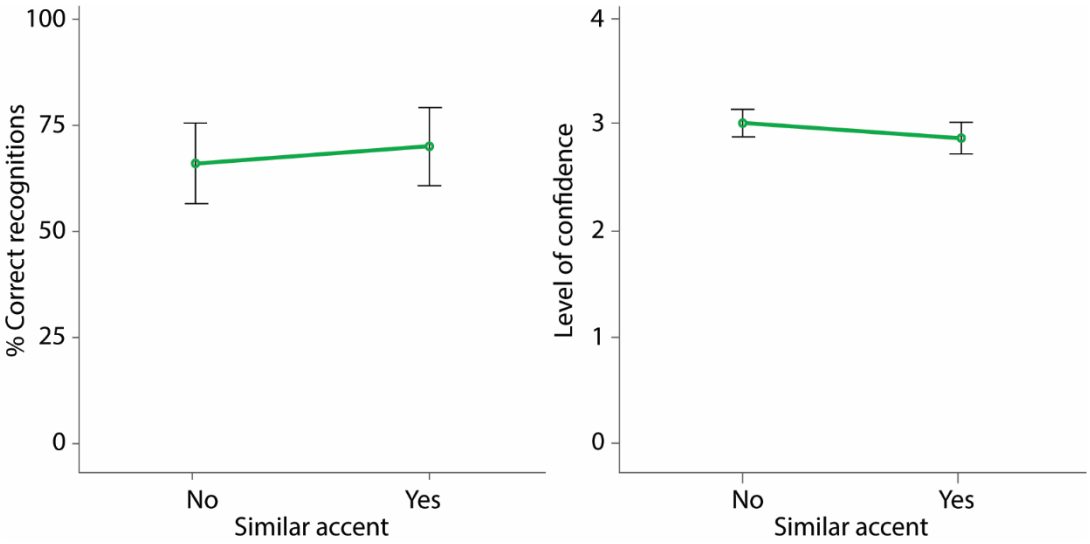

Figure 1. Successful recognition rate (left) and self-assessed confidence level (right) of participants in similar and different accent conditions. Standard errors are illustrated.

3.2 The OAE asymmetry

The second objective of the present paper was to investigate the asymmetry of this OAE with regard to the status of varieties of the same language. At first glance, Figure 2 (left) illustrates a tendency towards more accurate recognition for the Quebec participants in the similar accent condition, for whom the average successful recognition of the French target voice was 54% (SE 14.4 %), while that of the Quebec target voice was 64% (SE 13.3 %). On the other hand, the average number of successful recognitions by French participants was relatively similar for both varieties of French, i.e., 75% (SE 13.1 %) for the French target voice and 76% (SE 12.2 %) for the Quebec target voice. These data were analyzed using a generalized mixed-effects model with two fixed factors, participant origin and speaker origin, and participant identity as a random factor. The model's dependent variable was success on the recognition task. The model revealed no significant main effects or interactions. The absence of main effects indicates that there was no imbalance in the performance of the participants in each group according to their origin, nor according to the

origin of the speakers. In other words, none of the groups performed significantly better, and neither of the two line-ups was significantly easier. Again, the absence of a significant effect showing a more accurate recognition of participants' own accents may be due to the fact that the French people recruited had been living in Quebec for varying durations of time. It is possible that the OAE for people of French origin who have lived in Quebec for several years has decreased over time and exposure.

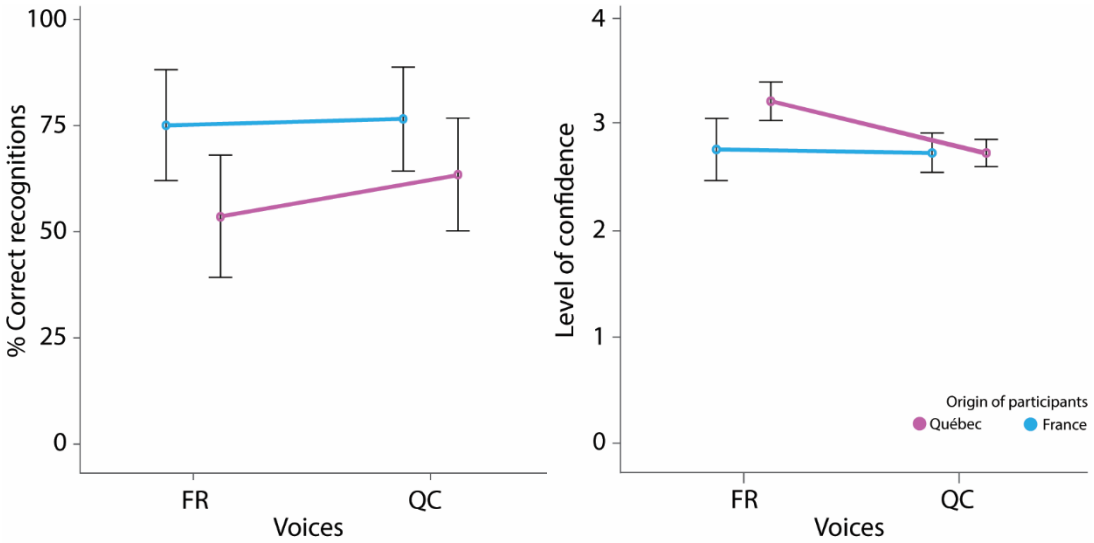

Figure 2. Successful recognition rate (left) and self-assessed confidence level (right) of participants from each of the French and Quebec origin groups for each of the French varieties studied. Standard errors are illustrated.

With regard to the self-assessed confidence level, the box on the right of Figure 2 illustrates a lesser average confidence level of Quebec participants when recognizing the Quebec target voice (2.93/4, SE 0.13). When recognizing the French target voice, the average was 3.08/4 (SE 0.18). However, their recognition performance was lower. In other words, the Quebec participants were more confident about the accuracy of their answers, but performed more poorly. The group of French participants had identical mean confidence levels for both conditions, i.e., 2.92/4 (SE = 0.29 for the FH line-up and 0.19 for the FQ line-up). Their confidence level was therefore unaffected by the origin of the speakers. Once again, it is possible to postulate that their establishment in Quebec may have had an impact on this result. A mixed model using the origin of participants (Quebec or France) and that of speakers (FQ and FH) as two fixed factors proved not significant.

3.3 Duration of stay

Considering that the French participants recruited for the present study had been living in Quebec for a variable number of years, the final question of interest here concerns the possibility of a difference in results depending on the time elapsed since their establishment. Given the non-significant results for the OAE, it is interesting to see whether an effect of duration of stay in Quebec (stay duration effect) can be observed for the French participants. Figure 3 (left) illustrates a gain in the percentage of correct recognitions of the target Quebec voice as the duration of stay increases. The percentage of recognition of the Quebec voice for participants established in Quebec for less than 5 years (n = 7) is 67% (SE 21.1%), while this rate rises to 86% (SE 14.3%) for people established in Quebec for 5 years or more (n = 9). Conversely, recognition of the French

target voice declines for participants who have been in Quebec longer (67%, SE 21.1%), while this percentage for participants with a shorter stay is 83% (SE 16.7%). To analyze this question, we used a generalized mixed-effects model with two fixed factors, namely the duration of participants' stay in Quebec (2 levels categorical variable) and the origin of the speakers, and the identity of the participants as a random factor. Here again, the model proved non-significant.

Finally, the last dependent variable to have been analyzed is the level of self-assessed confidence as a function of the duration of time participants from France have been living in Quebec. Figure 3 (right) shows that French participants gain in confidence over time with regard to FH. Their average confidence for the short stay duration is 2.67/4 (SE 0.33), while it is 3.17/4 (SE 0.23) for the long stay duration. However, this rising confidence is somewhat of an illusion, as their performance in recognizing the French target voice actually declines over time (Figure 3, left). A similar tendency can be seen in the confidence level regarding the Quebec target voice. Their average confidence level for short stay is 2.75/4 (SE 0.31), while long stay results in a confidence level of 3.07/4 (SE 0.47). A mixed model using the origin of the speakers (FQ and FH) and the duration of stay of French participants (short and long) as fixed factors was non-significant.

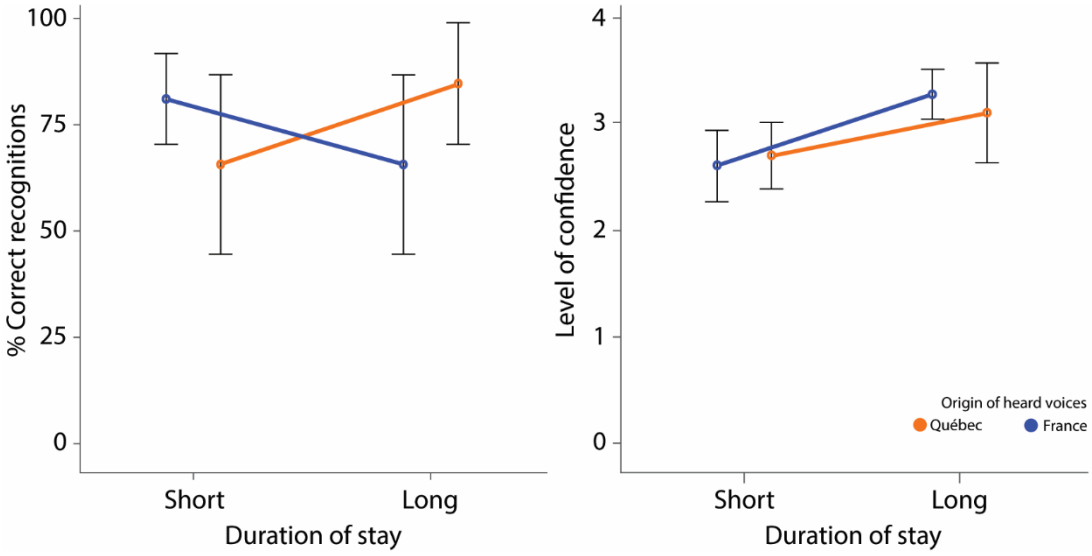

Figure 3. Comparison between the rate of successful recognition (left) and the average confidence scores of French participants (right), according to the duration of their stay in Quebec and the variety of French assessed. Standard errors are illustrated.

In short, although the results are non-significant, there is a tendency for French participants to perform better and be more confident in their FQ-related responses when they have been established in Quebec for a longer time. Conversely, their performance diminishes with regard to FH voices, but their confidence does not.

4 Discussion

The main objective of this study was to investigate whether the expertise effect, and more specifically the OAE, affected the accuracy of speaker recognition for FQ and FH. Secondarily, we also wanted to investigate whether this effect affected individuals from France and Quebec asymmetrically in terms of the dominance of one variant over the other. Finally, the effect of duration of stay in Quebec for French participants was explored, to determine whether longer exposure to an accent or given variety of language could attenuate the OAE. To answer these

questions, two voice line-ups were elaborated, one in FQ and the other in FH, and participants from Quebec and France were asked to identify to the best of their ability initially unknown voices to which they had only been briefly exposed.

4.1 Other-accent effect

According to the statistical analysis carried out, the population sample observed in the present study was not affected by the OAE in the speaker recognition task. Although a tendency confirming this effect seemed to emerge for the Quebec participants, it turned out to be non-significant. This statistical non-significance could, however, be explained by the fact that the French participants performed very similarly for both varieties of French. Yu et al. (2021) suggest four possible explanations to the OAE: 1. the phonologies of both language varieties is different, which leads to an increased difficulty in the task for the unfamiliar variety; 2. the social belonging to one of the linguistic groups leads to better performances for that group; 3. a lesser understanding of one of the varieties leads to impaired performances in this variety and 4. a greater number of exemplars of the linguistics schemes of one variety compared to the other would lead to asymmetrical performances (see Goggin et al., 1991). Considering the similarities between FQ and FH phonological systems (there are some differences, but they remain, ultimately, one language) and that the intelligibility of the texts used as stimuli was controlled for in both varieties, explanations 1 and 3 could account for the lack of significant effect. These explanations were also considered by both Yu et al. (2021) and Johnson et al. (2018). As mentioned in section 1.2, the language familiarity effect (LFE) has been largely attested whereas the OAE remains debated. Some studies have reported corroborating data (Atkinson, 2015; Stevenage et al., 2012; Vanags et al., 2005) and others, such as Yu et al. (2021) and Johnson et al. (2018), have only found an LFE but no OAE. For example, Johnson et al. (2018) reported that participants performed significantly better in a speaker recognition task in their native tongue (American English) compared to a foreign language, Dutch. They, however, showed no differences when tested in American or Australian English. In a similar way, Yu et al. (2021) reported significantly different results when Canadian participants were tested in Canadian English and English spoken with a strong Mandarin accent but not differences when comparing results between Canadian and Australian English. Going back to Yu et al. (2021) explanations 1 and 3, a possible account for the absence of OAE between the present FQ and FH voice line-ups and both participants groups could be that the phonological proximity between those two varieties is greater than, for example, that of English and Dutch or Canadian English and English with a strong Mandarin accent.

Furthermore, it is clear that the best way to test for an expertise effect would have been to ensure that the groups of individuals who took part in the experiment, either as speakers or as participants, remained more impervious to contact with the other variety of French used. However, like the participants from France, the speakers who took part in the recording of the voice samples had been established in Quebec for a number of years (between two and four). It is therefore possible that they incorporated certain audible elements of FQ (intonation, rhythm, etc.) into their speech. Consequently, in order to ensure that the recorded accent was representative of the hexagonal variety and devoid of contact with FQ, the speakers would have had to be recruited directly in France. It is also worth noting that targeting a single region of France, or even a single city (e.g., Paris) for sampling, could have greatly helped to ensure a more accurate match of accents. However, since the experiment was conducted in Montreal, the available speakers necessarily had experience with FQ because they lived in Montreal. Recruiting participants who had never lived in Quebec, or who had only lived in Quebec for a very short time, would also have

been interesting to some extent. However, as will be discussed in the conclusion, this approach is less interesting from a forensic phonetics point of view.

4.2 Expertise effect/asymmetry

We originally postulated that FQ participants would be less affected by the OAE. The "standard French" status that is generally attributed to FH throughout the French-speaking world makes it a more widely spread variety that Quebecers are accustomed to hearing. The tendency illustrated in Figure 2 (left) shows that Quebec participants may recognize the Quebec voice slightly better than the French voice, but French participants did not show the same tendency. The expected asymmetry thus appears to be a non-significant tendency only among Quebec participants. Our data supports explanations 2 and 4 of Yu et al. (2021). On one side, Quebec participants have a greater number of FQ exemplars compared to FH, but French participants, being established in Quebec for varying durations, could have stored similar amounts of exemplars for both varieties. As for the social belonging, it could have shifted from France to Quebec, for some at least, leading to similar performances in both conditions. We suggest that recently arrived French speakers may have shown an asymmetry in their responses while the time spent in contact with the Quebec accent may have neutralized it for those who have lived in Quebec longer. This might have resulted in the non-significant tendency observed in our results.

The similarity of the French participants' confidence results suggests that the origin of the speakers did not affect their perception of their ability to recognize the target voices. In other words, they were just as confident in identifying a Quebec voice as they were in identifying a voice of their own native variety, which is also consistent with hypotheses about the presence of the French in Quebec.

4.3 Duration of stay

The different durations the French participants had been established in Quebec allowed us to divide them into two separate groups ("less than 5 years" and "5 years or more") in order to examine more closely the effect of prolonged exposure to a different accent on speaker recognition. We observed a tendency towards better recognition of the Quebec target voice in the "5 years or more" category compared to the "less than 5 years." As for the French target voice, its successful recognition rate was higher when the duration of stay was short and tended to decrease with increasing years. Although these tendencies are not significant, they support the hypothesis that the OAE is present in the early years of a stay but fades as experience with the new accent increases. Explanations 2 and 4 from Yu et al. (2021) also support these observations. A more in-depth study using more data specifically focused on this question could confirm these tendencies. As mentioned above, several trials had to be eliminated due to the order in which the voices were presented, resulting in a smaller number of responses to analyze. Duration of stay was also a secondary issue in the present study, and the number of participants of French origin was therefore not based on this unique factor. In view of the non-significant tendencies reported herein, it would be reasonable to assume that a future study about the OAE focusing on people of French origin living in Quebec, with particular attention paid to the duration of their stay in Quebec, would yield conclusive results.

The results of the confidence analyses show that, somewhat predictably and in line with the hypotheses, French participants gained confidence in the Quebec target voice the longer they stayed in Quebec. Surprisingly, however, this increase in confidence also holds for the French target voice. The second possible explanation given by Yu et al. (2021) concerns the social belonging of participants to one of the linguistic groups involved. In the present situation, the gain

in confidence shown by the French participants does support this explanation. Regarding the growing confidence of French participants towards their performances with the French voices as their stay in Quebec becomes longer, we suggest that these participants grow a feeling of French expertise when comparing themselves to their environment (Quebec). The scope of this study and our data is insufficient to bring clear answers on this question, but it seems evident that further research investigating the effect of the duration of stay on the OAE is necessary.

5 Conclusion

In view of the results obtained in the present study, and despite the fact that most of the hypotheses were not statistically confirmed, we can nevertheless make some observations. On the one hand, as mentioned in the previous section, it goes without saying that recruiting participants living in France, who have had minimal contact with Quebecers, might have yielded more precise results regarding the OAE. However, from a forensic phonetics standpoint, the choice of participants based in Quebec is more interesting and more representative of the needs of the field. In fact, this article explores the question of the OAE with the aim of advancing forensic phonetics science in the Quebec context. There is very little reason to believe that a person who lives in France and who has had no contact with Quebec French could be asked to collaborate on a police investigation with jurisdiction in Quebec and involving only Quebec voices. On the other hand, it would be very plausible for a French individual living in Quebec to be approached for this purpose, and more specifically to identify a Quebec voice. Our choice of participants was therefore more appropriate and relevant to our objective despite its limitations. Following the results of our experiment, the variety of French would not have an effect on the performances on a speaker recognition task. A French witness could therefore testify during a legal procedure involving voice line-ups and speaker recognition. Given the tendencies observed in our results, combined with results reported by other scientists, we, however, recommend great caution if such a situation were to occur. As mentioned before, it seems that further research is required, especially regarding the French varieties and the impact of the duration of stay on the OAE.

On the other hand, it seems that speaker recognition is a complex process about which, to date, we do not have enough knowledge. It is undeniable that further study and investigation on the subject should be made in order to ensure that a recognition made in the context of a voice line-up can be considered irrefutable proof of guilt. As Atkinson (2015, p. 24) states, “It is not possible to cross-examine or provide expert analysis of a naive listener’s memory or identification of a speaker.” This makes it all the more important to have as detailed an understanding as possible of the reliability of speaker recognition. As far as the legal system is concerned, caution should therefore be exercised when considering testimony based on the recognition of a speaker for whom the witness has limited expertise.

Acknowledgements

We would like to thank Pr. Lucie Ménard for her support and access to the facilities and equipment of the Phonetics Laboratory at UQAM.

References

Atkinson, N. (2015). *Variable factors affecting voice identification in forensic contexts* [PhD thesis, University of York]. White Rose eTheses Online. https://etheses.whiterose.ac.uk/13013/

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell., S. (2021). On the dangers of stochastic parrots: Can language models be too big? 🦜. In *2021 ACM Conference on*

Fairness, Accountability, and Transparency (FAccT '21) (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Betancourt, K. S., & Bahr, R. H. (2010). The influence of signal complexity on speaker identification. *The International Journal of Speech, Language and the Law*, *17*(2), 179–200. https://doi.org/10.1558/ijsll.v17i2.179

Brasseur, A., & Ménard, L. (2013). Les marqueurs dialectaux du français québécois: Perception de locuteurs québécois. In *La perception des accents du français hors de France. Actes du colloque international d'Avignon, 17-18 novembre 2011* (pp. 103–127).

Broeders, A. P. A., & van Amelsvoort, A. G. (1999). Lineup construction for forensic earwitness identification: A practical approach. Communication presented at *14th International Congress of Phonetic Sciences*, San Francisco, CA.

Bull, R., & Clifford, B. R. (1999). 4. Earwitness testimony. *Medicine, Science and the Law*, *39*(2), 120–127. https://doi.org/10.1177/002580249903900206

Clifford, B. R., Rathborn, H., & Bull, R. (1981). The effects of delay on voice recognition accuracy. *Law and Human Behavior*, *5*, 201–208. https://doi.org/10.1007/BF01044763

Consulat général de France à Québec. (2020, April 8). *La communauté française au Québec*. https://quebec.consulfrance.org/La-communaute-francaise-au-Quebec

Cook, S., & Wilding, J. (1997). Earwitness testimony: Never mind the variety, hear the length. *Applied Cognitive Psychology*, *11*(2), 95–111. https://doi.org/10.1002/(SICI)1099-0720(199704)11:2<95::AID-ACP429>3.0.CO;2-O

de Jong-Lendle, G., Nolan, F., McDougall, K., & Hudson, T. (2015). Voice lineups: A practical guide. In *Proceedings of the International Congress of Phonetic Sciences* (pp. 10–14).

Doty, N. D. (1998). The influence of nationality on the accuracy of face and voice recognition. *American Journal of Psychology*, *111*(2), 191–214. https://doi.org/10.2307/1423486

Eriksson, E. J. (2007). *That voice sounds familiar: Factors in speaker recognition.* [Doctoral thesis, Department of Philosophy and Linguistics, Umeå University]. WorldCat.org. https://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-1106

Foulkes, P., & Barron, A. (2000). Telephone speaker recognition amongst members of a close social network. *International Journal of Speech, Language and the Law*, *7*(2), 180–198. https://doi.org/10.1558/sll.2000.7.2.180

Foulkes, P., & French, P. (2012). Forensic speaker comparison: A linguistic–acoustic perspective. In L. M. Solan & P. M. Tiersma (Eds.), *The Oxford handbook of language and law* (pp. 558–572). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199572120.013.0041

Fox, J., & Weisberg, S. (2020). car: Companion to Applied Regression. [R package]. https://cran.r-project.org/package=car.

Gallucci, M. (2019). GAMLj: General analyses for linear models. [jamovi module]. https://gamlj.github.io/.

Gastaldi, J. L. (2021). Why can computers understand natural language? *Philosophy & Technology*, *34*(1), 149–214. https://doi.org/10.1007/s13347-020-00393-9

Goggin, J. P., Thompson, C. P., Strube, G., & Simental, L. R. (1991). The role of language familiarity in voice identification. *Memory & Cognition*, *19*(5), 448–458. https://doi.org/10.3758/bf03199567

Goldstein, A. G., Knight, P., Bailis, K., & Conover, J. (1981). Recognition memory for accented and unaccented voices. *Bulletin of the Psychonomic Society*, *17*(5), 217–220. https://doi.org/10.3758/bf03333718

Harvey, M. B., Giroux, M., & Price, H. L. (2023). Lineup size influences voice identification accuracy. *Applied Cognitive Psychology*, *37*(6), 1303–1315. https://doi.org/10.1002/acp.4124

Hayward, W. G., Rhodes, G., & Schwaninger, A. (2008). An own-race advantage for components as well as configurations in face recognition. *Cognition*, *106*(2), 1017–1027. https://doi.org/10.1016/j.cognition.2007.04.002

Hollien, H. (2012). About forensic phonetics. *Linguistica*, *52*(1), 27–53. https://doi.org/10.4312/linguistica.52.1.27-53

The jamovi project. (2023). jamovi (Version 2.3) [Software]. https://www.jamovi.org

Jessen, M. (2008). Forensic phonetics. *Language and Linguistics Compass*, *2*(4), 671–711. https://doi.org/10.1111/j.1749-818X.2008.00066.x

Jessen, M. (2020). Speaker profiling and forensic voice comparison: The auditory-acoustic approach. In M. Coulthard, A. May, & R. Sousa-Silva (Eds.), *The Routledge handbook of forensic linguistics* (2nd ed., pp. 382–399). Routledge. https://doi.org/10.4324/9780429030581-31

Johnson, E. K., Bruggeman, L., & Cutler, A. (2018). Abstraction and the (misnamed) language familiarity effect. *Cognitive Science*, *42*(2), 633–645. https://doi.org/10.1111/cogs.12520

Kerstholt, J. H., Jansen, N. J. M., van Amelsvoort, A. G., & Broeders, A. P. A. (2006). Earwitnesses: Effects of accent, retention and telephone. *Applied Cognitive Psychology*, *20*(2), 187–197. https://doi.org/10.1002/acp.1175

Kircher, R. (2012). How pluricentric is the French language? An investigation of attitudes towards Quebec French compared to European French. *Journal of French Language Studies*, *22*(3), 345–370. https://doi.org/10.1017/S0959269512000014

Kreiman, J., & Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. John Wiley & Sons. https://doi.org/10.1002/9781444395068

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608. https://doi.org/10.1126/science.1736364

May, A., Sousa-Silva, R., & Coulthard, M. (2020). Introduction. In M. Coulthard, A. May, & R. Sousa-Silva (Eds.), *The Routledge handbook of forensic linguistics* (2nd ed., pp. 1–8). Routledge. https://doi.org/10.4324/9780429030581-1

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*(1), 3–35. https://doi.org/10.1037/1076-8971.7.1.3

McDougall, K. (2013). Assessing perceived voice similarity using Multidimensional Scaling for the construction of voice parades. *International Journal of Speech, Language and the Law*, *20*(2), 163–172. https://doi.org/10.1558/ijsll.v20i2.163

Nolan, F. (1991). Forensic phonetics. *Journal of Linguistics*, *27*(2), 483–493. https://doi.org/10.1017/S0022226700012755

Nolan, F. (2003). A recent voice parade. *International Journal of Speech, Language and the Law*, *10*(2), 277–291. https://doi.org/10.1558/sll.2003.10.2.277

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*, 195–203. https://dx.doi.org/10.3758/s13428-018-01193-y

Perrachione, T. K., Chiao, J. Y., & Wong, P. C. M. (2010). Asymmetric cultural effects on perceptual expertise underlie an own-race bias for voices. *Cognition, 114*(1), 42–55. https://doi.org/10.1016/j.cognition.2009.08.012

Perrachione, T. K. (2018). Speaker recognition across languages. In S. Frühholz & P. Belin (Eds.), *The Oxford handbook of voice perception*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198743187.013.23

Philippon, A. C., Cherryman, J., Bull, R., & Vrij, A. (2007). Earwitness identification performance: The effect of language, target, deliberate strategies and indirect measures. *Applied Cognitive Psychology*, *21*(4), 539–550. https://doi.org/10.1002/acp.1296

Plante-Hébert, J., & Boucher, V. J. (2014). *L'identification vocale : Pour une quantification des effets de la familiarité*. Communication presented at Journée d'Études sur la Parole, Le Mans. http://www.afcp-parole.org/doc/Archives_JEP/2014_XXXe_JEP_LeMans/2014_XXXe_JEP_LeMans.pdf

Plante-Hébert, J., & Boucher, V. J. (2015). *Effects of nasality and utterance length on the recognition of familiar speakers*. Communication presented at 18th International Congress of Phonetic Sciences, Glasgow. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0772.pdf

Plante-Hébert, J., Boucher, V. J., & Jemel, B. (2021). The processing of intimately familiar and unfamiliar voices: Specific neural responses of speaker recognition and identification. *PLoS ONE*, *16*(4), e0250214. https://doi.org/10.1371/journal.pone.0250214

Proulx, P., & Messier, L.-P. (2020). *Ces audacieuses qui ont façonné le Québec: 60 portraits de femmes entêtées.* Les Éditions du Journal.

Public Prosecution Service of Canada. (2018). *L'innocence en péril : La nécessité de vigilance continue afin de prévenir les condamnations injustifiées au Canada.* https://www.ppsc-sppc.gc.ca/fra/pub/ip-is/index.html

R Core Team (2021). R: A Language and environment for statistical computing. (Version 4.1) [Software]. https://cran.r-project.org. (Package retrieved January 1, 2022, from MRAN snapshot).

Reinke, K., & Ostiguy, L. (2012). Doublage et sociolinguistique: Une étude comparative du doublage québécois et français. *Zeitschrift für Kanada-Studien*, *32*(1), 26-48.

Rhodes, G., Hayward, W. G., & Winkler, C. (2006). Expert face coding: Configural and component coding of own-race and other-race faces. *Psychonomic Bulletin & Review*, *13*(3), 499–505. https://doi.org/10.3758/BF03193876

Robitaille, A. (2023). *Portrait sociodémographique et régional des personnes admises au Québec de 2011 à 2020 et présentes en 2022.* Ministère de l'Immigration, de la Francisation et de l'Intégration. Gouvernement du Québec.

Rose, P. (2002). *Forensic speaker identification* (1st ed.). CRC Press. https://doi.org/10.1201/9780203166369

Statistics Canada. (2022, August 22). *Police-reported crime statistics in Canada, 2021*. https://www150.statcan.gc.ca/n1/pub/85-002-x/2022001/article/00013-eng.htm#r9

Stevenage, S. V., Clarke, G., & McNeill, A. (2012). The "other-accent" effect in voice recognition. *Journal of Cognitive Psychology*, *24*(6), 647–653. https://doi.org/10.1080/20445911.2012.675321

Stevenage, S. V., & Neil, G. J. (2014). Hearing faces and seeing voices: The integration and interaction of face and voice processing. *Psychologica Belgica*, *54*(3), 266–281. http://dx.doi.org/10.5334/pb.ar

Thompson, C. P. (1987). A language effect in voice identification. *Applied Cognitive Psychology*, *1*(2), 121–131. https://doi.org/10.1002/acp.2350010205

Vanags, T., Carroll, M., & Perfect, T. J. (2005). Verbal overshadowing: a sound theory in voice recognition? *Applied Cognitive Psychology*, *19*(9), 1127–1144. https://doi.org/10.1002/acp.1160

Watt, D., & Brown, G. (2020). Forensic phonetics and automatic speaker recognition: The complementarity of human- and machine-based forensic speaker comparison. In M. Coulthard, A. May, & R. Sousa-Silva (Eds.), *The Routledge handbook of forensic linguistics* (2nd ed., pp. 400–415). Routledge. https://doi.org/10.4324/9780429030581-32

Wilding, J., Cook, S., & Davis, J. (2000). Sound familiar? *Psychologist*, *13*(11), 558–562.

Yu, M. E., Schertz, J., & Johnson, E. K. (2021). The other accent effect in talker recognition: Now you see it, now you don't. *Cognitive Science*, *45*(6), e12986. https://doi.org/10.1111/cogs.12986

Zarate, J. M., Tian, X., Woods, K. J. P., & Poeppel, D. (2015). Multiple levels of linguistic and paralinguistic features contribute to voice recognition. *Scientific Reports*, *5*(1), 11475. https://doi.org/10.1038/srep11475

Appendix A: Text read by speakers during the exposure phase

"La bise et le soleil se disputaient, chacun assurant qu'il était le plus fort, quand ils ont vu un voyageur qui s'avançait, enveloppé dans son manteau. Ils sont tombés d'accord que celui qui arriverait le premier à faire ôter son manteau au voyageur serait regardé comme le plus fort. Alors, la bise s'est mise à souffler de toute sa force mais plus elle soufflait, plus le voyageur serrait son manteau autour de lui et à la fin, la bise a renoncé à le lui faire ôter. Alors le soleil a commencé à briller et au bout d'un moment, le voyageur, réchauffé a ôté son manteau. Ainsi, la bise a dû reconnaître que le soleil était le plus fort des deux."

Appendix B: Text read by speakers during the testing phase

"J'ai pu constater moi-même l'absence des femmes dans notre histoire. On peut se demander si c'est un oubli volontaire de la part des historiens ou si ça ne reflète pas tout simplement la position inférieure dans laquelle la femme, dès les débuts de la colonie, était tenue." (Proulx & Messier, 2020)