THE EFFECT OF FOREIGN FILM ON THE PRODUCTION AND PERCEPTION OF NON-NATIVE SPEECH

by

Amy E. Hutchinson

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Linguistics

West Lafayette, Indiana

May 2022

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Olga Dmitrieva, Chair
School of Languages and Cultures

Dr. Alexander L. Francis
College of Health and Human Sciences

Dr. Daniel J. Olson
School of Languages and Cultures

Dr. Jessica L. Sturm
School of Languages and Cultures

Dr. Mariko M. Wei
School of Languages and Cultures

Approved by:
Dr. Olga Dmitrieva

ACKNOWLEDGMENTS

This dissertation would not have been possible without the support of many. First and foremost, I would like to thank my major professor, Dr. Olga Dmitrieva, for her guidance, encouragement, and patience. Olga's mentorship has been constant over the past six years, and I would not be the scholar I am today without her unending guidance and generous commitment to my growth as a researcher. I am also grateful to each my dissertation committee members, all of whom provided invaluable advice and expertise throughout this journey. I would like to thank Dr. Jessica Sturm for her support since day one of my graduate career and for inspiring the idea for this dissertation in her class on second language teaching with film. Similar gratitude is due to Dr. Dan Olson for his thoughtful and patient advice on various stages of the research process, as well as his general guidance on a range of topics related to academia. I am thankful to Dr. Alex Francis for his expertise in phonetics and experimental design, both of which were critical during the development of this project and others. Additionally, I owe gratitude to Dr. Mariko Wei for her kindness and for fueling my interest in individual variation in second language acquisition. Outside of my committee, I am thankful to Dr. Alejandro Cuza for his tremendous support during my time at Purdue.

I am additionally grateful to a number of people in the Purdue community who have provided support and invested their time in me throughout my graduate career. Thank you to my former (and forever) lab mate, Jenna Conklin, for everything from teaching me how to formulate IF statements in Excel my first year to driving me around southern Australia. Special thanks are also due to Adrien Bernardinis and Josh Weirick for providing their voices (literally!) for this dissertation. Continuing, I am thankful to Heather Offerman for guiding me through my first linear mixed effects models in R and to Juliette Duthoit for her help developing the French stimuli and instructions for various parts of this study. My utmost gratitude also goes out to Anthony Faulkner and Clara Villion who collected and annotated the majority of the data for this dissertation. Together Anthony and Clara put in hundreds of hours into this research, and this study would have either have never been possible or never been completed without them. This dissertation was primarily funded by the Linguistics Department at Purdue University, for which I am extremely grateful. I would also like to thank the Office of Interdisciplinary Graduate Programs for providing

me with the Bilsland Dissertation Fellowship, which allowed me to take a year to solely invest in my research. Finally, I am thankful to the almost 300 individuals who took the time to participate this study.

I am very fortunate to have friends both near and far who have supported me and provided me with continuous love and encouragement. I am grateful to all of my Purdue colleagues, many of whom have now become some of my closest friends. I have wonderful friends, all of whom have made this process a bit more bearable whether they realized it or not.

I would like to thank my parents for their unconditional love, encouragement, and support. My parents were the ones who nurtured my curiosity throughout my life and instilled in me a love of learning. Most of what I am today is owed to them. I am also thankful for my sister, the littlest nug, who has always made life a lot more fun. I am grateful to have shared the wild journey of graduate school with my partner, Josh Weirick, whose love and support I would never have made it through without. Finally, what kind of childless millennial would I be if I did not thank my cat, Buster Sailboat Hutchinson Weirick, who as I write these acknowledgements is snoozing on my lap? Throughout our time together, Buster has provided me with endless companionship and a lot of joy, so I dedicate this dissertation to him even though he does not know what a dissertation is.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The present dissertation explores the effect of exposure to non-native speech via foreign film on non-native speech production and perception. In order to explore potential effects, two main experiments were developed, which examined French production and perception by monolingual native speakers of English before and after exposure to French film. Across both experiments, two variables were selected for observation: high rounded vowels (/y/ and /u/) and consonant voicing (VOT). The production component of the dissertation investigated whether exposure to French film aided in the ability of monolingual American English speakers (n=74) to shadow French words containing high rounded vowels, /y/ and /u/, as tested through acoustic analyses and native French listener perceptual judgements (n=221). Perception of high rounded vowels and consonant voicing were examined using a perceptual assimilation task with category goodness ratings and a binary forced-choice voicing identification task, respectively.

With regard to the role of foreign film in non-native speech production, results indicated that a single session of exposure to French film had a small but significant effect on shadowing of French /y/, which was also perceptible to native French listeners. Shadowing of /u/, however, was not significantly affected by exposure. Additionally, while the acoustic analysis of VOT did not reveal any significant effects of film, native French listeners perceived post-film exposure productions to be significantly more target-like than pre-film exposure productions. This finding suggests that although VOT was seemingly unaffected by foreign film exposure, participants may have adjusted alternative acoustic correlates of voicing and that these modifications were perceptible to native listeners. In general, these results suggest that while potential effects of film are present, they are highly dependent on the variable being observed.

Results from the perceptual portion of the dissertation do not provide evidence that film exposure was effective at influencing non-native speech perception for either high rounded vowels or consonant voicing. However, it is suggested that this could be due to the difficulty of the tasks chosen rather than the effectiveness of foreign film.

Taken together, the present dissertation provides evidence that exposure to non-native speech via foreign film can affect some aspects of non-native speech learning. It is hypothesized

that further sessions may compound these initial benefits, especially in those who are already learning a second language.

CHAPTER 1. INTRODUCTION

1.1 Introduction and Goals

Acquisition of second language speech, both its production and comprehension, is an integral part of learning a second language (L2). Although a recent shift in the perspective on the goals of second language speech learning emphasizes intelligibility over native-like pronunciation (for review, see Levis, 2005), L2 pronunciation that approximates that of the target population of speakers remains one of the factors contributing to intelligibility of L2 speech (Derwing et al., 1998; Derwing & Munro, 1997, 2009). Research on pronunciation teaching and learning has been undergoing a revival in recent years, as both instructors and learners recognize it as a need often not met in a traditional second language classroom (Deng, 2009; Elliott, 1995; Levis & Grant, 2003; Lord, 2008; Offerman, 2014; Olson, 2014; Sturm, 2019; among others). With these broad directions in mind, the main question in the foreground of the present research concerns the types of exposure and training that can help learners improve their L2 pronunciation and comprehension in an efficient way.

Previous research on the acquisition of second language speech overwhelmingly demonstrates that exposure to authentic native speaker input is key for attaining target-like pronunciation and perception of non-native speech (Flege, 2007, 2009, 2018; Flege et al., 1997; Flege et al., 1995; MacKay et al., 2001; among others). For example, there is evidence that learners immersed in an L2 environment, for instance study abroad programs, outperform classroom learners with respect to pronunciation (Freed et al., 2004; O'Brien, 2004; Segalowitz et al., 2004; Stevens, 2002, 2011). While immersion clearly aids in the development of L2 speech, this option is not always accessible for all learners and it is worth exploring how other types of naturalistic exposure can contribute to the development of L2 speech. The current study explores this broad question by examining how an alternative form of authentic native input, specifically, foreign film, can contribute to the acquisition of non-native speech.

In domains other than speech, foreign film has been demonstrated to effectively broaden L2 vocabulary, while also developing communicative, cultural, and symbolic competence (Dubreil, 2011; Herron et al., 1999, 2000; Kaiser, 2011; Kaiser & Shibahara, 2014; Sherman, 2006; Sturm, 2012; Tognozzi, 2010; Zhang, 2011). Additionally, foreign film provides viewers with rich, highly

variable input consisting of a variety of tokens produced by multiple talkers, which previous research has demonstrated to be an effective way to improve non-native speech production and perception (Bradlow, 2008; Bradlow et al., 1997, 1999; Logan et al., 1991; Pisoni et al., 1982; among others). Finally, previous literature suggests that synchronized audio-visual input (i.e. the ability to see speakers' face while listening) leads to more successful speech perception and information retention than audio input alone (Hardison, 2003, 2005; Hazan et al., 2005; Hirata & Kelly, 2010; Inceoglu, 2016; among others). Exposure to foreign film provides a combination of these benefits, potentially compounding their positive effects. It is therefore hypothesized that film, although not specifically developed for the task of improving non-native pronunciation and perception, will significantly promote the acquisition of non-native speech.

The present dissertation investigates the effect of foreign film on non-native speech acquisition by examining French production and perception by three groups (two experimental and one control) of monolingual, native English participants (n= 74) before and after exposure to French film. Two specific variables were selected for analysis: high rounded vowels /y/ and /u/, and consonant voicing (VOT). The production component of the study consisted of a shadowing task whereby participants were asked to repeat a French word containing one of the aforementioned variables following a native French model talker. Their performance on this task was analyzed acoustically, as well as through judgements provided by native French listeners (n= 225) via an AXB perceptual judgement task (Pardo et al., 2010). Perception of high rounded vowels and consonant voicing were examined using a perceptual assimilation task with category goodness ratings (Best, 1995) and a binary forced-choice voicing identification task (used to establish a perceptual boundary), respectively.

Confirming the effect that L2 film has on the production and perception of L2 speech would be pivotal in the field of second language acquisition. In a world that is becoming increasingly globalized, multimedia streaming services like Netflix and Hulu make it significantly easier to access audiovisual content in other languages, providing a low-cost and potentially effective alternative to interactions with native speakers. Studies demonstrating the effectiveness of foreign film in improving non-native speech are a critical first step towards incorporating it more consistently into the second language classroom.

1.2 Outline of the Dissertation

This dissertation consists of eight chapters, including the present introduction chapter. Chapter 2 provides a background on topics relevant to the current dissertation, including non-native speech training procedures and previous training with foreign film, as well as non-native speech production and perception of high rounded vowels and VOT. This chapter also presents a review of the production/perception relationship in L2 speech and introduces the research questions and hypotheses.

Chapter 3 presents a broad view of the methodology of the present study, providing information about the participants and general procedures.

Chapter 4 presents a more detailed overview of methods used in the shadowing portion of the study, providing specific information about the stimuli, procedures, data processing, and analysis. It also reports the results and provides a brief interim discussion. This chapter is divided into two sub-sections based on the variable in question: high rounded vowels or VOT.

Chapters 5 and 6 follow a similar structure to Chapter 4 (methodology, results, interim discussion), presenting a comprehensive view of the native listener perceptual judgement task and the perception tasks, respectively.

Chapter 7 provides a general discussion of all of the results from Chapters 4-6, connecting the different experiments and providing a more global interpretation of the results.

The dissertation is concluded with Chapter 8 that presents a summary of the results and discusses its limitations and future directions.

CHAPTER 2. LITERATURE REVIEW

The present chapter presents a detailed review of the literature relevant to the current dissertation. Section 2.1 details previous training programs developed in order to improve non-native speech production and perception, specifically focusing on two forms of training that are particularly relevant to the study: high variability phonetic training (section 2.1.1) and audiovisual training (section 2.1.2). The section concludes with a brief outline of previous research on training with film, as well as examining the potential effects of viewing a film with subtitles (section 2.1.3).

Continuing, section 2.2 presents a review of previous literature on non-native speech production, specifically focusing on production of non-native high rounded vowels (section 2.3.1) and VOT (section 2.3.2). In addition to outlining previous production research, section 2.2 also provides information on the acoustics of the variables in question. Section 2.3 mirrors the preceding section by providing a review of previous literature on the perception of target variables in non-native speech: non-native high rounded vowels (section 2.4.1) and VOT (section 2.4.2). The literature review portion of the chapter concludes with a review of the production/perception relationship in L2 speech (section 2.5). To conclude, section 2.6 details the research questions and hypotheses the dissertation aims to address.

2.1 Training of non-native speech

When acquiring a second language, non-native learners must "re-tune" their sound system from the structure of their native language to a system that can flexibly switch back and forth from the native to the non-native language(s) (Bradlow, 2008; Iverson et al., 2003). While the development of this sound system flexibility is quite challenging for non-native speakers, previous research has demonstrated the effectiveness of training programs in easing some of the challenge for both non-native speech perception (Bradlow et al., 1999; Flege et al., 1996; Hazan et al., 2005; Iverson et al., 2003; Jamieson & Morosan, 1986; Kazanina et al., 2006; Kewley-Port & Watson, 1994; Nishi & Kewley-Port, 2005; Nishi Kanae & Kewley-Port, 2007, 2008; Pisoni et al., 1982; Strange & Dittmann, 1984; Wang et al., 1999; among others) and production (Akahane-Yamada et al., 1998; Bradlow et al., 1997, 1999; Catford & Pisoni, 1970; Dalby et al., 1998; Dalby & Kewley-Port,

1999, 1999; Flege et al., 1995; Hazan et al., 2005; Hirata & Kelly, 2010a; Inceoglu, 2016; among others).

One of the first studies to apply a training program in order to aid in the perception of non-native phoneme contrasts was Strange & Dittmann (1984). The goal of the program was to aid native speakers of Japanese in the development of the English /ɹ-l/ contrast. In this seminal study, participants were trained to discriminate between two sounds by listening to them on a synthetic, *rock-lock*, speech continuum. Exposure to this continuum allowed participants to home in on the acoustic parameters that differentiate the two sounds and, following training, participants were significantly better at discriminating between the two as determined by discrimination performance on a novel /ɹ-l/ continuum (*rake-lake*). The positive findings from this study encouraged a plethora of future research that trained non-native speakers to produce and perceive a variety of contrasts by employing various training programs on a variety of variables. The subsections that follow will outline previous literature on two particularly effective forms of non-native training approaches: high variability phonetic training (HVPT) and audiovisual training. The focus on these specific forms of training in previous research is because it is suggested that film exposure also can provide highly variable audio-visual input, potentially offering similar benefits.

2.1.1 High variability phonetic training

Training procedures aid in the development of L2 speech production and perception likely because they mimic critical aspects of natural exposure through immersion. A substantial portion of this research highlights the effectiveness of HVPT approach (Bradlow, 2008b; Bradlow et al., 1997, 1999; Lively et al., 1993, 1994; Logan et al., 1991), a naturalistic procedure that emphasizes exposure to audio produced by many different talkers across a variety of diverse tokens. That is, while Strange & Dittmann (1984) trained participants using single minimal pair (*rock-lock*) on a synthetic continuum (representative of a single "talker"), HVPT research supports the idea that the more talkers and tokens a non-native speaker is presented with, the more successful the training program will be. The first study to apply such a program, Logan et al. (1991), expanded the low-variability training program created by Strange & Dittmann (1984) by exposing participants to a wider range of stimuli along an /ɹ-l/ continuum naturally produced by a variety of American

English talkers. More naturalistic exposure led to a significant improvement in minimal pair perceptual identification for all talkers in the post-test phase of the experiment. Lively et al. (1994) expanded these findings by providing evidence that HVPT training is effective at improving perception of English /ɹ-l/ by monolingual Japanese speakers (not just current English learners as examined Logan et al., 1991), and that improvement was retained 6 months after the initial training phase despite participants not receiving any additional training.

Following HVPT's success in the perception domain, Bradlow et al. (1997) found that perceptual gains made during training were also transferable to the production domain when analyzing how Japanese speakers post-training productions of English /ɹ/ and /l/ were perceived by native English listeners. Further research determined that the effect of HVPT on production was also apparent in delayed post-test 3 months after training (Bradlow et al., 1999).

Since its inception, may researchers have adopted an HVPT approach and it has been successfully used to train a number of other variables like production and perception of Mandarin lexical tone by native English speakers (Wang et al., 1999, 2003; Wiener et al., 2020), categorization of non-low German vowels by native English speakers (Kingston, 2003), perception and production of English /e/ and /æ/ by Cantonese learners of English (Wong, 2012), identification of Hindi stops (dental and retroflex) by native speakers of Japanese and English (Pruitt et al., 2006), and voice identity (Lavan et al., 2019), among others. HVPT training has also been demonstrated to be effective at training a number of different populations, including different ages (for example, Giannakopoulou et al., 2017 examined both adults and 8 year old children), language pairings, and settings (for example, Thomson, 2011, 2012 and Thomson et al., 2016 all applied the program in a classroom setting). All of these examples demonstrate the relative success and flexibility of the HVPT paradigm, suggesting its ability to aid in non-native speech development across a variety of variables, settings, and populations.

2.1.2 Audiovisual training

A considerable amount of HVPT research has taken training one step forward by supplementing the audio signal with a visual component (often a human or synthetic face that is synchronized to audio stimuli). A large body of research has demonstrated the superiority of audiovisual training when compared to audio-only, likely due to the McGurk effect that demonstrates that audio and

visual information is integrated during speech perception (McGurk & MacDonald, 1976). However, existing research on audiovisual training has mostly been limited to the perception of consonant contrasts (Hardison, 1999, 2003; Hazan et al., 2005, 2006; Lively et al., 1993; Ortega-Llebaria et al., 2001; Shen, 2019; Wang et al., 2008, 2009; among others). For example, when examining the effect of auditory vs. audiovisual perceptual training on the perception of the English /v/-/b/-/p/ contrast, as well as the perception and production of the English /l/-/r/ contrast by native Japanese learners of English, Hazan et al. (2005) found that participants, in general, were positively influenced by the addition of a visual modality. With the exception of /l/-/r/ contrast perception, perception and production was significantly superior in learners who received audiovisual training (/l/-/r/ production significantly improved following audiovisual training even though perception was unaffected), especially if a human face, as opposed to a synthetic one, was used.

In general, research on the effect of audiovisual training on speech production and/or perception appear to support the addition of a visual component. However, results are inconclusive in the areas that have generally received less attention, like vowel perception/production. One study that attempted to fill these gaps, Inceoglu (2016), presented research on the effect of two types of training, audiovisual and audio-only, on the perception and production of French nasal vowels by current French learners (English as a first language, L1). Results indicated that production improvement was significantly greater for talkers in the audiovisual group compared to those in the audio-only group. However, although both groups significantly improved in perceptual performance after training, they did not significantly differ based on the type of information they received (audio and/or visual).

One notable difference between the results of the two previously mentioned studies (Hazan et al., 2005 and Inceoglu, 2016) are the differences in the success of participants who receive an additional visual component. Hazan et al. (2005) found significant differences between the majority of contrasts and modalities trained when comparing auditory and audiovisual training, while Inceoglu (2016), did not find the modalities to perform significantly differently with regard to perceptual performance. In this case, it is likely that the difference is present due to the phonological contrast examined (Inceoglu, 2016). Nasal vowels (studied by Inceoglu, 2016), while visually salient to native speakers of French, are not as obvious to non-native speakers, while consonants are more visually distinct. Thus, one conclusion that arises from the contrasting results

of previous research is that not all non-native speech categories are likely to benefit equally from audiovisual training.

In general, while evidence on the benefits of audiovisual training appears to be mixed, it appears that it is relatively successful, although the degree of success is modulated by the visual characteristics of the contrasts. Nevertheless, as evidenced by Hirata & Kelly (2010), additional modalities on top of audio and visual have the potential to do so. When examining the effect of four types of training (audio-only, audio-mouth, audio-hands, or audio-mouth-hands) on the ability of monolingual native speakers of American English to perceive Japanese vowel length contrasts, Hirata & Kelly (2010) demonstrated that participants who received mouth movement input significantly outperformed those who only received the audio modality. Contrastively, treatment that included hand gestures was not effective at improving identification. In the latter case, it is suggested that more input does not always equate to better learning and that, in some cases, it can overload participants cognitively, hindering acquisition.

2.1.3 Training with film

While film combines aspects of HVPT (i.e. exposure to natural productions by multiple talkers and a wide range of tokens) with the benefits of both an audio and visual component, it has not previously been investigated with regard to its potential influence on non-native speech development. In realms other than speech, research has demonstrated foreign film's ability to broaden L2 vocabulary (Kaiser, 2011; Sherman, 2006), as well as aid in the development of communicative (Sherman, 2006), cultural (Dubreil, 2011; Herron et al., 1999, 2000; Sturm, 2012; Tognozzi, 2010; Zhang, 2011), and symbolic competence (Kaiser & Shibahara, 2014).

Additional questions that arise with regard to the potential of film to improve non-native speech production and perception concern the use of subtitles. That is, whether foreign film is a more effective tool when watched with subtitles (either in the viewer's first language or in the original language of the film) or without (audio-only). Literature on the language of subtitles generally finds that subtitles in the L2 are helpful to viewers, while native language subtitles can hinder film effectiveness in areas like speech perception (Mitterer & McQueen, 2009; Vanderplank, 2016). However, for some L2 learners and non-native speakers, L1 subtitles are the only option as they are not competent enough in the film's language to get any benefit from foreign subtitles.

This raises the question of whether native language subtitles lead to more benefit than simply watching a film without them. Some research supports the idea that reading subtitles requires more cognitive resources and prioritizes visual input over audio (Danan, 1992), which might reduce the potential benefit of viewing the film. However, more recent research on the effect of subtitling on second language pronunciation found captioned video to significantly lead to pronunciation gains (Wisniewska & Mora, 2020). One factor not addressed in the previous research is that depending on how proficient the viewer is in the language of the film, not having access to subtitles could likely cause them to lose attention to the film out of boredom. With this in mind, further research on the effect of subtitling should be completed in order to determine to what level subtitles are affective at positively influencing L2 acquisition.

To summarize, foreign film in the original language arguably combines the benefits of both high variability speech input and audiovisual exposure, while also being a highly accessible and engaging form of non-native speech exposure. Thus, exploring whether exposure to authentic film material in the target language aids in L2 learning, acquisition of non-native speech specifically, is a goal worth pursuing, as it can help determine whether incorporating a film component into the L2 classroom can be a relatively easy and effective way to boost speech learning. The present dissertation takes the first steps towards exploring the possible benefits of film exposure for L2 speech learning. Specifically, it investigates whether a relatively brief (45 minute), single session of exposure to French film material improves the perception and production of French high rounded vowels and consonant voicing (VOT) in naïve native monolingual speakers of American English.

The present study specifically examines naïve learners in order to control the amount of previous exposure to the target language. French rounded vowels were chosen because they are among the more challenging French speech sound categories shown to be difficult for native English speakers to acquire (Flege, 1987; Flege & Hillenbrand, 1984; Gottfried, 1984; Levy, 2009a, 2009b; Levy & Law, 2008, 2010; Levy & Strange, 2008; Rochet, 1995; Strange et al., 2009). Moreover, lip rounding is a visually salient feature of vowel production; thus, the acquisition of rounded vowels could benefit from exposure to visual information in film (Traunmüller & Öhrström, 2007). Consonant voicing was examined due to the challenge that learners face when trying to acquire the French voicing distinction (Birdsong, 2007; Flege & Hillenbrand, 1984; Gabriel et al., 2016; Louër, forthcoming; Netelenbos et al., 2016). However, unlike rounded

vowels, voicing is not as visually salient. Finally, while a single brief exposure to film is likely to result only in subtle changes in participants' non-native speech production and perception patterns, the evidence of positive effects in such challenging conditions would provide a promising direction to explore in designs with more extensive exposure and more experienced participants.

2.2 Non-native speech learning

2.2.1 Theories of non-native speech acquisition

Considering the complexity of non-native speech development, many theories have been developed in order to account for the phenomena. These theories attempt to predict the way in which non-native speech will be produced and perceived based on the underlying knowledge of the learner. Two models that formalize speech learning and are particularly influential are the Speech Learning Model (SLM) (Flege, 1987, 1995, 2002; Flege & Bohn, 2021) and the Perceptual Assimilation Model (PAM) (Best, 1995). While both models make predictions about the acquisition of new phonetic categories, the PAM only predicts L2 perceptual learning, while the SLM attempts to account for both L2 perception and production. Additionally, the SLM addresses L2 speech learning, while the PAM accounts for the perception of non-native contrasts by naïve listeners (though see the PAM-L2 for an extension of the theory to SLA; Best & Tyler, 2007) In particular, the SLM assumes that the mechanism of equivalence classification plays an important role at the beginning stages of the learning process. Due to equivalence classification non-native sounds are perceptually mapped to perceptually similar existing speech categories in the learner's L1. Equivalence classification can help or hinder acquisition of non-native speech sounds based on how similar they are to existing categories. The more perceptually similar a non-native, new sound is to the learners L1, the more likely the learner is to overlook the subtle but important phonetic differences and equate the two categories. As a result of equivalence classification, formation of separate phonological categories, necessary for the accurate production and perception of L2 sounds, can be blocked. For this reason, similar L2 categories are ultimately predicted to be more challenging for learners to acquire than sounds that do not resemble sounds that exist in the L1.

The SLM accounts for a number of phenomena that occur during speech acquisition and predicts potential challenges that might occur during the learning process. For example, an English

learner acquiring French may have a relatively easy time producing and perceiving French /y/ because a similar vowel does not exist in the English vowel inventory and, therefore, the French vowel is classified, according to the SLM, as a "new" sound. On the other hand, sounds like French /u/ and /t/, which are quite similar to their English counterparts, may pose challenges for English learners of French, as they are likely to equate the new sounds to their existing English category, blocking L2 category formation.

Like the SLM, the PAM proposes that not all phonetic categories will be perceived the same by non-native listeners, however, unlike the SLM, the PAM does not rely on categories as mental representations, but rather focuses on detection of information from articulatory gestures of the speech signal using a direct realist perspective (Best, 1995). In this case, an L2 phoneme's articulatory gestures are perceptually assimilated based on the articulatory gestures of a related phoneme in the native language. If the articulatory gestures of two L2 phonemes are perceived as being particularly distinct and they correspond to a distinct phoneme in the L1, there is two-category assimilation, and it is predicted that discrimination of the two phonemes will be very good. If one L2 phoneme is a better representative of an L1 phoneme than another L2 phoneme, this creates category goodness assimilation and assimilation is predicted to be intermediate. Finally, the PAM predicts that if there are two L2 phonemes that share the same gestures as one L1 phoneme, this results in single category assimilation and discrimination of the two L2 phonemes is hypothesized to be poor. Thus, like the SLM, the PAM creates a scale of perceptual discrimination difficulty based on how a sound is assimilated to a single L1 sound.

It is critical to consider theoretical predictions of models like the SLM and the PAM when attempting to account for non-native speech acquisition, as they can predict which phonetic categories will be more challenging for learners, and why.

2.3 Non-native speech production

2.3.1 Production of non-native high rounded vowels

Both high rounded vowels in French, /y/ and /u/, differ acoustically from their closest English counterparts, /i/ and /u/, but not to the same degree (see Figure 2.1 for English and French vowel spaces, as reported by Strange et al., 2007). French vowel /y/, which does not have a direct equivalent in English, is a high front rounded vowel (Battye et al., 2003; Fougeron & Smith, 1999;

Tranel & Bernard, 1987). It is acoustically the closest to English high front unrounded tense vowel /i/, differing primarily in terms of the second formant frequency: French /y/ has an appreciably lower F2 due to lip rounding, which lengthens the oral cavity and has a pronounced lowering effect on second formant frequency (Levy, 2004). In contrast, back rounded vowel /u/ can be found in both French and English. However, French /u/ has a somewhat lower F1 and F2, resulting in a higher and more retracted vowel than English /u/ (Levy & Law, 2010; Strange et al., 2007, 2009).

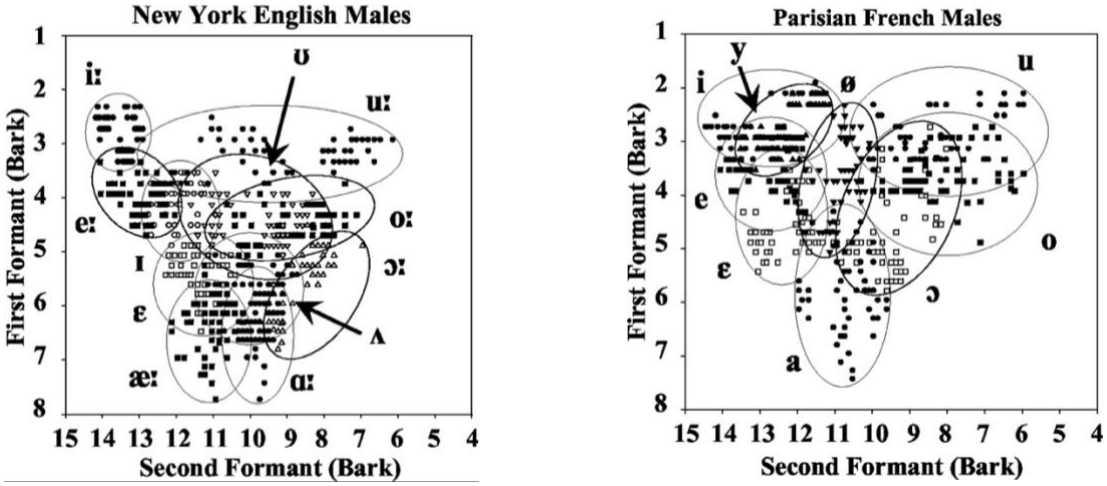

Figure 2.1. Vowel inventories of English and French as produced by native speakers of the languages (figures and data from Strange et al., 2007, p. 1116).

Given the challenges of high rounded vowel acquisition by native English speakers, ample research has been conducted on the matter. Previous research on the production and perception of rounded vowels with particular emphasis, where possible, on populations similar to our participants (naïve talkers) is outlined below.

Previous work has addressed how vowels /y/ and /u/ are produced by current L2 learners (English L1) acquiring languages like French or German (Flege, 1987; Flege & Hillenbrand, 1984; Levy & Law, 2008, 2010; Rochet, 1995). In general, /y/ has been found to present fewer difficulties for target-like production than /u/, possibly because it is novel to L1 English speakers. Vowel /u/ on the other hand is subject to a greater influence from a comparable L1 category, resulting in higher F2 values (a more front vowel) than would be expected in native French or German. For example, in a study that examined the effect of L2 experience on the production of French /y/ and

/u/ by native English speakers (Flege, 1987), all speakers, independent of experience level, produced French /y/ with near native-like formant values, while F2 values for /u/ were significantly higher than that of native speakers (similar results were reported in Flege & Hillenbrand, 1984). When interpreting these findings, it was suggested that phones like /u/, for which an acoustically similar counterpart exists in the speakers' L1, are more challenging to produce in a target-like manner, due to the equivalence classification with the native phone. In contrast, "new" phones like /y/, for which a close equivalent is not found in the talkers' L1, are easier to categorize as different from L1 phones and are easier to produce more accurately. Findings like these supported the development of the SLM (Flege 1987, 1995, 2002; Flege & Bohn, 2021), which, as previously outlined, attributes the challenges of L2 phone production to their acoustic and perceptual similarity to existing L1 counterparts.

Of particular interest to the current study is previous literature that examines the production of French /y/ and /u/ by speakers with no prior experience with the language. While this research is limited, it generally supports the predictions of SLM. For example, Levy & Law (2010) found that naïve talkers were able to create a front-back distinction between French /y/ and /u/ in an alveolar context, producing /y/ with better accuracy than /u/ (although in bilabial context talkers produced French /u/ more accurately than /y/).

2.3.2 Production of non-native VOT

The second variable in question in the current study is VOT, which is defined as the period of time between a stop consonants release and the onset of vocal cord vibration (Lisker & Abramson, 1964). As voicing is realized differently in French and English, acoustic differences make production challenging for learners. Native speakers of French, a 'voice language' (Iverson & Salmons, 2003), realize voiced stops in word-initial position with prevoicing (negative VOT < 0 ms) and voiceless stops with short lag (positive VOT < 30 ms) (Caramazza & Yeni-Komshian, 1974). Comparatively, native speakers of English, an 'aspiration language' (Iverson & Salmons, 2003), realize word-initial voice stops primarily with short lag and voiceless stops with long lag (positive VOT > 30 ms) (Lisker & Abramson, 1964, 1967). Word-initial prevoicing is also possible in English as a sub-phonemic variant of the voiced category (Davidson, 2016; Flege, 1982; Keating, 1984; Lisker & Abramson, 1964, 1967; Westbury, 1979). As demonstrated in Table 2.1,

these categorical distinctions, in particular the use of short lag in different phonological categories by the two languages (voiceless in French and voiced in English), make acquisition challenging for the L2 learner. As previous research on variation in VOT produced by native and non-native speakers has determined that it can contribute to foreign accent and potential miscommunication (Gonzalez-Bueno, 1997; Schoonmaker-Gates, 2015), it is important for non-native speakers to learn to differentiate the realization of voicing across languages.

Table 2.1. Realization of voicing in French and English.

	[+voice]	**[-voice]**
French	Prevoiced *(negative VOT < 0 ms)*	Short lag *(positive VOT < 30 ms)*
English	Short lag (primarily); prevoicing (sub-phonemic variant)	Long lag *(positive VOT >30 ms)*

As the realization of voicing poses a potential challenge during L2 acquisition and is relatively easy to measure, a wide array of research has been conducted in order to investigate how the contrast is acquired by native English speakers. Past research primarily explores how current L2 learners or bilinguals acquire the contrast, mainly focusing on English learners of 'voice languages' like French or Spanish or vice versa (Spanish/French L1 and English L2) (Birdsong, 2007; Casillas, 2020; Flege, 1987; Flege & Eefting, 1988; Flege & Hillenbrand, 1984; Gabriel et al., 2016; González López & Counselman, 2013; Lord, 2008; Louër, forthcoming; Nagle, 2017; Netelenbos et al., 2016; Reeder, 1998; Schuhmann & Huffman, 2019; Zampini, 1998). In general, previous research suggests that voicing acquisition is heavily influenced by the L1 and subject to transfer effects, especially by relatively inexperienced learners. For example, Flege (1987) compared voiceless stop acquisition across three groups of French learners (L1 English) and found that only the most experienced group (American expatriates who had resided in Paris for an average of 11.70 years) produced target-like VOT. The least experienced group, college students who had spent 9 months in Paris 3-6 months prior to the study, completely relied on L1 transfer, and the mid-experience group, university-level instructors of French, produced intermediate values that fell between English and French. Results from studies that report voiced stop acquisition support similar findings, where target-like production is highly dependent on level of experience

with the L2 (Casillas, 2020a; Gabriel et al., 2016; Louër, forthcoming; Nagle, 2017; Yang et al., 2020).

Similar to rounded vowels, research focused on how monolingual speakers of English produce non-native voicing contrasts is relatively limited. Flege & Eefting (1988) examined the imitation of a VOT continuum (10 ms steps from -60 ms to 90 ms) by monolingual native English adults and found that the majority of stimuli that fell between -60 ms and 30 ms were produced with short lag VOT despite the presence of prevoicing in some tokens. Monolingual English adults also primarily imitated items that fell between 50 ms and 90 ms with long lag. While the latter finding is not particularly surprising, this study is primarily useful in explaining how new phonetic categories (i.e. prevoicing) are challenging for speakers of English even when presented in an imitation format.

Another study that examined how monolingual English speakers produce both extended and reduced synthetic voiceless VOTs (Nielsen, 2011), is relevant when attempting to explain how naïve talkers might produce durations that differ from native-like norms. In this case, a shadowing task revealed that talkers were able to significantly increase voiceless durations but not shorten them. This latter finding is particularly relevant in the current study, as French voiceless stops are realized with short lag and talkers must shorten voiceless VOTs in order to achieve target-like production.

2.4 Non-native speech perception

2.4.1 Perception of non-native high rounded vowels

The first variable that the perceptual portion of the study investigates is the perception of non-native (in this case, French) high rounded vowels, /y/ and /u/. As outlined in section 2.3.1, these two vowels pose different degrees of challenge for English learners of languages where the contrast exists (i.e. French, German, etc.) due to their existence, or lack thereof, in the English vowel inventory. French vowel /y/ (high front rounded) does not exist in the English vowel inventory and while /u/ does, it is slightly higher and more back when compared to English, which makes target-like production and perception challenging. Of particular interest with regard to the present variables is that despite French /y/ being more acoustically similar to English /i/ than /u/, French /y/ and /u/ often show single-category assimilation to English category (/u/) making discrimination

of the pair challenging (Best, 1995; Best & Tyler, 2007). That said, previous research on the perception of /y/ and /u/ indicates that while English speakers do not always have difficulty distinguishing between the two vowels, assimilating the vowels to native categories suggests more confusion. For example, in a simple AX discrimination task, native English listeners could differentiate between French /y/ and /u/ with near native-like accuracy, regardless of the level of second language proficiency (Best et al., 1996; Flege & Hillenbrand, 1984; Polka, 1995). More complex, AXB-style discrimination tasks, however, presented greater challenges for listeners across L2 experience levels, who frequently confused /y/ and /u/ (Gottfried, 1984; Levy, 2009b; Levy & Strange, 2008). Finally, research that deployed perceptual assimilation tasks revealed some support for single-category assimilation (Best, 1995) and therefore potential for confusion between /y/ and /u/. For example, in Strange et al. (2009), 11 naïve American listeners heard French and German vowels, categorized them as the most similar sounding English vowels, and rated the goodness of fit to the selected English category. French back rounded vowel /u/ was assimilated to English /u/ 84% of the time with a high goodness of fit rating (7/9). French /y/ was most often assimilated to a back vowel (84% of the time) as well, despite its acoustic similarity to English front vowel /i/, suggesting perceptual salience of the rounding feature. Importantly, /y/ was categorized as English /u/ 52% of the time, albeit with a low goodness of fit ratings (2/9), revealing partly similar assimilation patterns for /y/ and /u/. Comparable findings were also reported by Levy (2009a) who found that a group of monolingual American English speakers categorized both French /y/ and /u/ as back vowels (/^{j}u/ and /u/, respectively), and by Rochet (1995) who reported similar results with a population of monolingual Canadian English speakers.

2.4.2 Perception of VOT

The second variable selected for perceptual analysis in the present experiment is VOT, as the temporal measure differs from English to French. In English, previous research suggests that the boundary between voiced and voiceless stops is located around 30 ms (Lisker & Abramson, 1970). That is, all stop consonants produced by talkers with VOTs less than ~30 ms are perceived by listeners as voiced and all consonants produced with VOTs greater than ~30 ms are perceived to be voiced. With regard to French, previous research suggests that, in general, the voicing boundary is around 0 ms (Caramazza & Yeni-Komshian, 1974; Laguitton, 1997; Serniclaes, 1987). That

said, some additional research has contested this boundary, with some experimental research identifying a boundary at around 7 ms (Medina et al., 2010) to as high as approximately 20 ms (Nearey & Rochet, 1994)[1]. Research has indicated that contrasting findings are due to the place of articulation of the French stop (Fischer-Jørgensen, 1972; Nearey & Rochet, 1994; Rochet et al., 1987), vowel context (Fischer-Jørgensen, 1972; Munro, 1987; Nearey & Rochet, 1994; Rochet et al., 1987), and dialect of French. While the exact boundary in both languages is subject to slight variation as function of rate of speech, place of articulation, and vowel context, under comparable experimental conditions English voicing boundary is expected to occur 10-30 ms later on the VOT continuum than the French one.

With regard to L2 acquisition, English learners of 'voice' languages (Iverson & Salmons, 2003) like French, must learn to adjust their voicing boundary in order to adequately communicate in the language, and several studies have investigated how learners, particularly French/English bilinguals, perceive the voicing distinction (Caramazza et al., 1973; Hazan & Boulakia, 1993; Mack, 1989). In general, research in this area suggests that bilinguals tend to crossover from voiced to voiceless at a point that is intermediate between French and English, though the exact crossover point is heavily dependent on age of L2 acquisition (Hazan & Boulakia, 1993; Mack, 1989). Input in the target language also appears to play a role in boundary placement, as demonstrated by Casillas (2020) who reported perceptual findings from L2 learners of Spanish, another 'voice' language, who were participating in a seven-week immersion program. In this case, by the end of the immersion program, participants' perceptual boundaries in L2 were nearly identical to native speakers.

While a decent amount is already known about the acquisition of voice onset time perception by English learners of French and other 'voice' languages, there appears to be a gap in the literature with regard to how voicing boundaries in unfamiliar languages are perceived by monolingual naïve English speakers. While non-native VOT boundaries have been examined in a training contexts (for example, Casillas, 2020b; Rochet & Chen, 1992) and baseline measurements might have the potential to provide a look at how monolingual speakers percieve non-native voicing contrasts, many of these studies used synthetic stimuli rather than resynthesized stimuli

[1] While Nearey & Rochet (1994) report a relatively high boundary, they claim that this is possibly due to a methodological error on their end. In this case, they presented participants with a 10 ms stimulus and a 29 ms stimulus, creating a 19 ms step size between the two critical items. The lack of stimulus items between 10 and 29 ms could have led to an overall higher voicing boundary for pariticpants.

from native speakers of the target language. With that in mind, many of the baseline measurements provided in previous literature are not entirely relevant to the present study, as it is highly likely that participants were simply treating the task like an examination of their L1. For example, Rochet & Chen (1992) investigated how native speakers of Mandarin [2] altered French perceptual boundaries following a three hour training program. Prior to treatment, participants percieved synthetic stimuli with relatively "Mandarin-like" boundaries, with crossover points ranging from around 30ms to 50ms. That said, as previously suggested, this finding should be interpreted with caution, as the stimuli used by Rochet & Chen (1992) were synthetic and not resynthesized French stimuli from native French speakers. In his analysis of the effects of a Spanish immerison program on perception of Spanish VOT by "absolute beginner" native English learners of Spanish, Casillas (2020) relied on resynthesized VOT tokens by a native Spanish speaker. Casillas (2020) reported Spanish VOT boundaries in the pre-test to be around 7 ms, demonstrating that natural stimuli likely have an effect on VOT boundaries regardless of whether the participant has previous experience in the target language.

2.5 Production/perception relationship

As the present dissertation examines both the perception and production domains, it is important to consider how the two are related and if one is more malleable to training than the other. In general, studies that examine both production and perception of non-native speech suggest that there is, at minimum, a modest correlation between the two (Bohn & Flege, 1992; Casserly & Pisoni, 2010; Evans & Alshangiti, 2018; Flege et al., 1997, 1999; Inceoglu, 2019; Levy & Law, 2010; Nagle, 2021; Rochet, 1995; Saito & van Poeteren, 2018), however, research is mixed on which realm of L2 speech (production or perception) precedes the other. It is generally suggested that perception must precede production (Levy & Law, 2010), as the same phenomena happens in L1 acquisition. That said, there is ample research suggesting that the opposite is true. For example, in their analysis of production and perception of English /ɹ/ and /l/ by Japanese learners of English, Sheldon & Strange (1982) found that participants had an easier time producing the contrast than perceiving it in an identification task. Flege & Eefting (1987) also found similar results when analyzing stop production and perception by Dutch learners of English. The learners' perceptual

[2] Mandarin is an 'aspiration' language (Iverson & Salmons, 2003) that realizes voicing in a similar manner to English.

boundary was only slightly shifted from Dutch to English, but, in production, they were able to create an ample acoustic difference in VOT duration between the two languages, ultimately suggesting that for those learners, mastery in production preceded perception (for other instances of production preceding perception, see Caramazza et al., 1973; Flege et al., 1997; Flege & Schmidt, 1995; Kleber et al., 2012; Mack, 1989; Schertz et al., 2015). Some of the previously reviewed training research also suggests that production might be mastered before perception (both, Hazan et al., 2005 and Inceoglu, 2016 for example, found audiovisual training improved all aspects of production but not of perception).

2.6 Research questions and hypotheses

The current dissertation presents two experiments that examine the role foreign film exposure in the production and perception of non-native speech. Based on reviewed literature, it is hypothesized that film will affect both production and perception, as it combines audio and visual modalities, which has previously proven to be a beneficial aspect of training (Hardison, 1999, 2003; Hazan et al., 2005, 2006; Inceoglu, 2016; Lively et al., 1993; Ortega-Llebaria et al., 2001; Shen, 2019; Wang et al., 2008, 2009). It is also noteworthy that foreign film exposes participants to real human faces, which have been demonstrated in past research (Hazan et al., 2005) to be more effective than exposure to a synthetic face. Additionally, foreign film exposes viewers to variation in talkers and tokens, which may be similar to that of effective training programs that use HVPT approaches (Bradlow, 2008b; Bradlow et al., 1997, 1999; Giannakopoulou et al., 2017; Kingston, 2003; Lavan et al., 2019; Lively et al., 1993, 1994; Logan et al., 1991; Pruitt et al., 2006; Thomson, 2011, 2012; Thomson et al., 2016; Wang et al., 1999, 2003; Wiener et al., 2020; Wong, 2012). Finally, though the film used in the present study was not specifically designed to improve speech production/perception, it provides participants with rich, authentic native-speaker input in another language, which has been demonstrated to be a necessity in the attainment of target-like pronunciation and auditory recognition of non-native speech (Flege, 2007, 2018; Flege & Liu, 2001; MacKay et al., 2001).

With consideration of the previously reviewed literature, the current dissertation presents two experiments that examine the role of foreign film exposure in the production and perception of French rounded vowels, /y/ and /u/, and consonant voicing (VOT). The first experiment

(Chapter 3) examines a potential effect of foreign film by presenting an acoustic analysis of the two variables in question. In addition to an acoustic analysis, Chapter 4 examines the role of foreign film in non-native speech production as determined by perceptual judgements provided by native French listeners. Finally, a perceptual assimilation task and a voicing identification (VOT boundary) task are implemented in Chapter 5 in order to determine how foreign film exposure influences rounded vowel and consonant perception, respectively.

While the dissertation broadly asks how exposure to foreign film effects non-native speech production and perception, the following research questions and hypotheses set more specific goals:

(RQ1) Will naïve, monolingual American English speakers shadow word medial French high rounded vowels (/y/ and /u/) in a more target-like manner after exposure to foreign film?

Hypothesis 1: Rounded vowels are hypothesized to be particularly susceptible to influence of film exposure as they provide the talker with additional, beneficial visual information (lip rounding) in addition to the acoustic signal (Traunmüller & Öhrström, 2007). This hypothesis is also supported by literature that demonstrates that vowel production can be improved with audiovisual training procedures (Inceoglu, 2016).

Additionally, previous research on the production of French /y/ and /u/ by native English talkers with varying amounts of French experience, demonstrates that "new" vowels like /y/ are easier to produce than "similar" vowels like /u/ (Flege, 1987; Flege & Hillenbrand, 1984; Levy & Law, 2008, 2010; Rochet, 1995). With that in mind, it is likely that talkers will improve more at /y/ than /u/ shadowing following exposure to foreign film (Flege, 1995, 2002). This research question and hypothesis is addressed in Chapter 4.

(RQ2) Will naïve monolingual American English speakers shadow French voicing of bilabial stops in word-initial position ([+voice] prevoiced, [-voice] short lag) in a more target-like manner after exposure to foreign film?

Hypothesis 2: In comparison to rounded vowels, improvement in VOT production is likely to be more challenging to talkers due to the lack of comparable visual saliency (Inceoglu, 2016). That said, it is hypothesized that talkers will demonstrate some changes in VOT duration and category frequency following exposure to foreign film, though these changes may be small and talkers are likely to rely heavily on L1 transfer during both pre- and post-

test sessions (Flege & Eefting, 1988; Nielsen, 2011). This research question and hypothesis is addressed in Chapter 4.

(RQ3) Will acoustic modifications to rounded vowels and consonant voicing made by monolingual American English talkers following foreign film exposure be perceptible to native French listeners?

Hypothesis 3: If talkers make substantial acoustic changes to their speech following exposure to foreign film, it is hypothesized that these changes will be perceptible to native French listeners. It is also hypothesized that since native listener perceptual judgements provide a more global assessment of phonetic convergence (Pardo et al., 2010), listeners may also pick up on modifications not analyzed in the acoustic analysis. This research question and hypothesis is addressed in Chapter 5.

(RQ4) Do non-native perceptual assimilation and category goodness ratings of French /y/ and /u/ change following exposure to foreign film?

Hypothesis 4: In line with previous literature that applied the same task type (Strange et al., 2009) to a similar population, it is hypothesized that initially all groups will most frequently categorize both French /y/ and /u/ as English /u/. Despite assimilating the two vowels to a single English category, it is predicted that French /u/ will be considered a relatively good fit for English /u/, as demonstrated with category goodness ratings, while /y/ will be considered a poor exemplar of English /u/. Following exposure to foreign film categorization of French /u/ is unlikely to change, however, it is predicted that participants will provide it with lower category goodness ratings after being made implicitly aware of the acoustic differences between the two vowels. With regard to French /y/, while film exposure might affect response categorization, as reflected by higher rates of assimilation to English /i/, ratings are likely to be affected the most. It is hypothesized that exposure to French film will cause participants to become more uncomfortable with any English response options, which will be reflected in category goodness ratings. This research question and hypothesis is addressed in Chapter 6.

(RQ5) Do non-native participants alter French voicing boundaries following exposure to foreign film?

Hypothesis 5: Casillas (2020) found that over the course of a seven-week immersion program, absolute beginner English learners of Spanish shifted their perceptual boundary in a target-like direction, likely due to a combination of exposure to the target language and explicit instruction. While participants in the current study will not have access to explicit instruction, it is hypothesized that exposure to native input via foreign film will cause them to slightly adjust their perceptual boundary to become more "French-like." This will be demonstrated by a more frequent selection of 'voiceless' that would result in a leftward shift of French perceptual boundaries. This research question and hypothesis is addressed in Chapter 6.

(RQ6) Does exposure to foreign film affect one realm of L2 speech (i.e production or perception) more than another?

Hypothesis 6: In line with the general assumption that speech perception precedes production (Levy & Law, 2010), it is hypothesized that talkers will make greater changes in perception following film exposure when compared to production. This research question and hypothesis is addressed in Chapter 7.

(RQ7) Will access to subtitling during foreign film exposure significantly affect non-native production/perception?

Hypothesis 7: Although both experimental groups in the current study will receive audiovisual components of the film, it is possible that the group viewing the film with English subtitles (Subtitled group) will focus exclusively on the subtitles instead of the film's visuals and thus will not have the benefit of attending to facial gestures. This is further supported by previous research on L2 film, which suggests that reading subtitles requires more cognitive resources and prioritizes visual input over audio (Danan, 1992). As watching a film in a language that talkers are not familiar already has the potential to cognitively overload participants, it is possible that the additional modality might hinder task outcomes (similar to the inclusion of a hand modality in Hirata & Kelly, 2010). With the previous research in mind, it is hypothesized that the experimental group that doesn't receive subtitles will out-perform those who do. This research question and hypothesis is addressed in Chapter 7.

CHAPTER 3. GENERAL METHODOLOGY

In order to address the research questions presented in Chapter 2, this chapter presents a general methodology of dissertation. While each of the following chapters also feature methodology sections that include more detailed information about each experiment/task (stimuli, procedures, data processing, and analysis), the current chapter details information that is relevant across all experiments. Section 3.1 presents details regarding the participants recruited for the current study. Moving forward, section 3.2 outlines the general procedures followed during the study, including information about film exposure, group assignment, and all pre/post-testing questionnaires.

3.1 Talkers

74 native speakers[3] (15 male, 57 female, 1 non-binary, and 1 declined to answer; mean age 21.17 y.o, SD=3.27) of American English with no prior French experience participated in the study. Recruitment took place on Purdue University's campus. All talkers indicated that they were raised in English-speaking homes[4] and spent the majority of their lives in the Midwest region of the United States. None of the talkers reported history of a speech or hearing disorder. Talkers were compensated $20 for their participation.

While talkers indicated no prior experience with the French language, many had some experience learning other languages, the most common of which was Spanish. However, only talkers who indicated relatively low proficiency in their L2 were recruited. When asked to rate their proficiency in other languages using a sliding Likert scale (1-7, 1 being "Very Poor" and 7 begin "Native-like"), the average was 2.37 (highest rating was 4).

Prior to participating in the study, talkers were randomly assigned to one of three groups: the Audio-Only group (n=25), the Subtitled group (n=24), or the Control group (n=15). Conditions for each group are outlined in section 3.2 of this chapter

Information regarding the native French listeners who participated in the AXB task of the study will be outlined in Chapter 5.

[3] 75 talkers were originally recruited to participate in the study, but one talker's data was excluded because they indicated previous experience learning French.

[4] Ten participants stated that their parents were non-native speakers of English but did not indicate that they had significant knowledge of the languages spoken by their parents.

3.2 Film exposure procedures

Prior to their lab visit, talkers completed a language background questionnaire (Appendix A) online through Qualtrics (Qualtrics, Provo, UT). This questionnaire was administered online in order to ensure that potential participants met the qualifications for the study (native speaker of English, no prior French experience), while also cutting down the time they spent in the lab. Once talkers arrived at the lab the study proceeded in the following order: Pre-test, Film exposure, and Post-test, followed by a post-participation questionnaire. Testing phases (pre- and post-test) were identical in structure for each talker but were counterbalanced between talkers. The shadowing task always preceded the perception tasks for all talkers so that participants did not have additional French speech exposure. The order in which talkers completed the perception tasks (VOT boundary task and perceptual assimilation task) was counterbalanced. The pre- and post-testing phases took talkers approximately 30-40 minutes to complete.

After completing the pre-test phase, talkers watched Season 1, Episode 1 of Chef's Table: France (Gelb, 2016), a documentary style show that reports on the day-to-day activities, motivations, and background of a French chef. This show was selected due to its short runtime (45 minutes and 23 seconds total), emotionally neutral topic, and visual nature (Sturm, 2012). Additionally, each episode of Chef's Table: France is self-contained, making it easy for viewers to follow without having seen previous episodes. Episode 1, which is focused on chef Alain Passard, was specifically chosen because his restaurant primarily cooks with vegetables and, unlike other episodes, there were very few scenes that involved meat, which might have disturbed some viewers. Talkers watched the episode at a computer station in a quiet room equipped with a sound dampening tri-fold stall and listened using a pair of Sennheiser HD 380 pro headphones.

Despite all talkers viewing the same episode during film exposure, the conditions of viewing depended on group assignment. Both experimental groups (Audio-Only and Subtitled) watched the episode completely in French. The Audio-Only group did not have access to any subtitles while watching the film, while the Subtitled group watched the film with English subtitles. This distinction was created in order to determine if access to meaning via subtitling hindered or facilitated learning of non-native speech. In order to verify that any potential effect found in either of the experimental groups was not due to a practice effect, a Control group was also included. Participants in the Control group watched the same episode, but instead of French audio the

episode was dubbed over in English. It is noted in the Control group, the French audio was not completely eliminated, but was included at a very low dB in the background.

To ensure that participants paid attention to the audio, they were asked to complete a simple while-watching activity, which was based on their group assignment (Appendix B and C) The two experimental groups (Audio-Only and Subtitled) were instructed to write French words that they recognized onto a sheet of paper. They were told that these could be French words they recognized (for example, *oui* or *magnifique*) or French words that sounded similar to English words (for example, *tarte* or *superb*). This specific activity was selected as it would encourage both groups to actively listen to the speech of the characters, rather than simply focus on the visuals, subtitles (when applicable), etc. The control group completed a similar task, but instead of French words, they recorded English food-related words (for example, *lobster* or *knife*). All groups were informed that their performance on the while-watching activity would not affect their compensation but were instructed to record as many words as they could. On average, the Audio-Only group recorded 38.30 words (SD= 22.82 words), the Subtitled group recorded 46.03 words (SD= 31.82 words), and the Control group recorded 79.47 words (SD=34.73 words).

Following film intervention and the post-test, all talkers completed a short background questionnaire on Qualtrics, which asked them to provide their opinions about the episode (Appendix D). This debriefing questionnaire was included in order to allow participants a place to record their thoughts on the episode/experiment, as well as to collect information about episode understanding and attention. Based on the questionnaire, only four talkers had previously seen an episode of Chef's Table, one of whom had specifically seen Chef's Table: France. On average, talkers indicated that they enjoyed the episode, rating it a 5.39 on a 7-point scale (1- disliked a great deal, 7-liked a great deal) (SD= 1.37). Average group ratings to additional questions are provided in Table 3.1. From start to finish, participants spent approximately 2 hours completing all of the experimental components.

Table 3.1. Average ratings to relevant post-participation questionnaire questions divided by group (Audio-Only, Subtitled, and Control) and average ratings across all groups (Combined). All questions ratings were based on a 7-point Likert scale and, in general, the high end of the scale represented a positive opinion.

	Audio-Only	**Subtitled**	**Control**	**Combined**
How much did you like watching the episode of Chef's Table: France today?	5.07	5.57	5.60	5.39
How much do you think you paid attention to the episode of Chef's Table: France today?	5.96	6.04	6.07	6.01
How much do you think you paid attention to the episode's audio (i.e. people talking, music, etc.)	5.58	5.92	5.33	5.70
Before and after watching Chef's Table, you completed a task where you repeated French words after a speaker. How effective do you believe watching Chef's Table was in improving your pronunciation of French words?	4.38	5.00	4.00	4.54

Table 3.2. Average ratings to relevant post-participation questionnaire questions divided by group (Audio-Only, Subtitled, and Control) and average ratings across all groups (Combined). All questions ratings were based on a 7-point Likert scale and, in general, the high end of the scale represented a positive opinion.

	Audio-Only	**Subtitled**	**Control**	**Combined**
How much did you like watching the episode of Chef's Table: France today?	5.07	5.57	5.60	5.39
How much do you think you paid attention to the episode of Chef's Table: France today?	5.96	6.04	6.07	6.01
How much do you think you paid attention to the episode's audio (i.e. people talking, music, etc.)	5.58	5.92	5.33	5.70
Before and after watching Chef's Table, you completed a task where you repeated French words after a speaker. How effective do you believe watching Chef's Table was in improving your pronunciation of French words?	4.38	5.00	4.00	4.54

CHAPTER 4. EFFECT OF FILM EXPOSURE ON SPEECH SHADOWING

As outlined in detail in Chapter 2, the goal of the current dissertation is to present a thorough analysis of the effect of exposure to foreign film on the production and perception of non-native speech. The current chapter is focused on production, specifically investigating how monolingual speakers of English shadow non-native French speech following exposure to French film. To do so, two variables were selected for analysis: French high rounded vowels, /y/ and /u/, and voice onset time of French stops /p/ and /b/. These particular variables were selected as they commonly pose challenges for English learners of French, due to their acoustic and articulatory attributes (rounded vowels: (Flege, 1987; Flege & Hillenbrand, 1984; Levy & Law, 2008, 2010; Rochet, 1995; VOT: Birdsong, 2007; Caramazza et al., 1973; Flege, 1987; Flege & Hillenbrand, 1984; Gabriel et al., 2016; Lord, 2008; Louër, forthcoming; Netelenbos et al., 2016). The present chapter is divided as follows: Section 4.1 provides a detailed methodology of the shadowing task, including information on stimuli (section 4.1.2), specific task procedures (section 4.1.3), and data processing and analysis of rounded vowels (section 4.1.4) and VOT (section 4.1.5). Section 4.2 presents the results which are divided by variable. Results from shadowing stimuli containing rounded vowels are presented in section 4.2.1 and an analysis of VOT duration and category frequency can be found in section 4.2.2. The chapter concludes with an interim discussion specific to shadowing results (section 4.3). The discussion section is divided into three sub-sections: the first one discusses the effect of film on rounded vowel shadowing (section 4.3.1), the second one discusses the effect of film on VOT shadowing (section 4.3.2), and the last which presents a more general discussion of film effects on non-native speech shadowing (section 4.3.3).

4.1 Methods

4.1.1 Talkers

All 74 talkers discussed in Chapter 3 (section 3.1) completed the shadowing task.

4.1.2 Stimuli

As the participants had no prior experience with French, a shadowing task was implemented (i.e. participants heard an item produced by a native speaker of French and repeated it back into a microphone). Model stimuli used for the shadowing task were produced by a 22-year-old, male native speaker of French. He was born in the Isère department in the Auvergne-Rhône-Alpes region in eastern France and spent the entirety of his life in that area. Despite being recorded in the United States, the model talker had only arrived in the country a week and a half prior to the recording session and had not previously spent time in English-speaking countries. Nevertheless, he was highly proficient in English and reported some knowledge of Spanish and Dutch.

Stimuli for the task were recorded by the model talker in a sound-attenuated booth using an ART Tube MP Project Series preamplifier and a Shure KSM32 Embossed Single-Diaphragm Microphone. Audio was captured using Audacity version 2.3.2 at a 44.1 sampling rate. The model talker repeated each isolated stimulus item three times and the second repetition was extracted. In instances where there was an issue with the second repetition (sound quality, etc.), the third repetition was extracted. All stimuli were extracted and normalized for intensity using Praat (Boersma & Weenink, 2018). Each stimulus consisted of a monosyllabic CV or CVC French word containing at least one of the target sounds: rounded vowels (/y/ or /u/) or word-initial stops (/p/, /b/). With the exception of the target-sounds, stimulus items did not contain sounds that were not also found in the English phonemic inventory (for example, /ʁ/)[5]. Where possible, words that contained multiple target sounds were used (for example, *puce* [pys] 'flea') in order to reduce the amount of time participants spent on the shadowing task. Each target sound was represented across 6 stimulus items resulting in 17 total items. A full list of stimuli, IPA transcriptions, and English translations are presented in Table 4.1 (/y/ or /u/) and Table 4.2 (/p/ or /b/). Items marked with an asterisk (*) indicate an item that appears on both lists. Stimuli were not controlled for frequency, as all of the words were novel to the monolingual English talkers in the present study.

[5] One exception to this case was the use of nasal vowel /ɛ̃/ in the bilabial stop stimuli *pain* (/pɛ̃/) and *bain* (/bɛ̃). Despite not existing in the English vowel inventory, /ɛ̃/ was included as nasality was originally intended to be analyzed. These items were ultimately left in the analysis as their exclusion did not significantly alter results.

Table 4.1. Stimuli, IPA transcriptions, and English translations of items containing French rounded vowels, /y/ or /u/.

Stimulus items containing /y/			**Stimulus items containing /u/**		
Item	Transcription	English Translation	Item	Transcription	English Translation
puce*	/pys/	flea	pouce*	/pus/	thumb
bulle*	/byl/	bubble	boule*	/bul/	ball
bu*	/by/	to drink (past tense)	bout*	/bu/	tip
vu	/vy/	to see (past tense)	toute	/tut/	all
pull*	/pyl/	sweater	cou	/ku/	neck
tu	/ty/	you (pronoun)	goût	/gu/	taste

Table 4.2. Stimuli, IPA transcriptions, and English translations of items containing French bilabial stops, /p/ or /b/.

Stimulus items containing /p/			**Stimulus items containing /b/**		
Item	Transcription	English Translation	Item	Transcription	English Translation
pêche	/pɛʃ/	peach	bêche	/bɛʃ/	spade
pain	/pɛ̃/	bread	bain	/bɛ̃/	bath
pull*	/pyl/	sweater	bulle*	/byl/	bubble
pouce*	/pus/	thumb	bout*	/bu/	tip
puce*	/pys/	flea	boule*	/bul/	ball
peine	/peine/	punishment	bu*	/by/	to drink (past tense)

The list of stimuli also contained nine additional items that were not analyzed, which for the purposes of this study served as fillers[6]. The additional items were identical in structure to the

[6] In the design phase of this dissertation, an analysis of the exposure of French film on vowel nasality was also planned. However, due to the complex acoustic nature of nasality and the tendency for participants to shadow the nasal vowels in the study as other sounds (e.g. rhotics, liquids, etc.) the nasal analysis ultimately was not carried out. However, the items that were originally designed for the nasal analysis (in Table 4.3) were left in the experiment.

stimuli but did not contain rounded vowels or bilabial stop consonants. A list of the additional items, their IPA transcriptions and English translations is provided in Table 4.3.

Table 4.3. IPA transcription and English translation of additional items.

Item	Transcription	English Translation
vin	/vɛ̃/	wine
lin	/lɛ̃/	flax (plant)
faim	/fɛ̃/	hunger
main	/mɛ̃/	hand
veine	/vɛn/	vein
laine	/lɛn/	wool
tonne	/tɔn/	metric ton
canne	/kan/	stick, walking stick
faune	/fon/	fauna

4.1.3 Procedures

During the pre- and post-test phases of the experiment, participants listened to the stimulus items produced by the model talker via headphones and were asked to repeat them into a microphone. When listening to the model talker's production of a French word, participants were presented with an animated picture of an ear on a red screen to indicate the listening portion of the trial. The screen changed to green with an animated mouth, 500 ms after the stimulus presentation [7], to indicate the speaking portion of the trial. Participants were given 2,500 ms to repeat the presented word into the microphone. This repetition was followed by a 500 ms blank screen and the next trial initiated. Each stimulus was presented for shadowing three time per session, in random order (6 repetitions total across the pre- and post-test).

[7] A 500 ms interstimulus interval was chosen as it encourages gradient processing rather than phonological processing, which could be affected by confounding variables like individual differences in phonological short term memory (Nagle & Baese-Berk, 2021; Werker & Logan, 1985).

Prior to the pre- and post-test trials, talkers were asked to read aloud a short text (Appendix E) in order to calibrate the microphones and allow talkers to become familiar with the procedures, sound booth, microphones, etc. Following the short text, talkers completed a practice trial that asked them to imitate English words that had been produced by the same model talker they would imitate during the trials. The practice trial consisted of six English words (*boat, boot, debt, den, peak, and pet*) and was identical in structure to the experimental trials outlined above. A practice trail was included so that talkers could familiarize themselves with the voice of the model talker and the structure of the trials.

All data collection took place in a sound-attenuated room and the experiment, as well as all pre-trial components, was presented to talkers using PsychoPy software (Peirce et al., 2019). Speech was recorded using an ART Tube MP Project Series preamplifier and a Shure KSM32 Embossed Single-Diaphragm Microphone. Audio was collected using Audacity at a 44.1 sampling rate and tokens were played for each participant through a pair of Sennheiser HD 380 pro headphones. Each recording session took approximately 7 minutes to complete.

4.1.4 Data processing and analysis of rounded vowels

All experimental items were annotated manually in Praat (Boersma & Weenink, 2018). Vowels were identified using the onset and offset of periodicity, which aided in ascertaining the transition between consonants and vowels. Praat's Burg LPC-based algorithm was used to extract the first, second, and third formant frequency at the midpoint of each vowel. In the event of errors in the LPC-based formants, measurements were hand-corrected by a researcher using the formants visible on the spectrogram. Formants were then transformed to Barks using the PhonR package (McCloy, 2012) in R version 4.1 (R Core Team, 2019) in order to make comparisons between speakers.

Statistical analysis was performed using the lme4 package (Bates et al., 2015) in R. To determine if film exposure influenced rounded vowel production, Bark-normalized F1 and F2 values were submitted as dependent variables to separate linear mixed effects (LME) models with Group (Audio-Only, Subtitled, and Control) and Session (pre-test and post-test) as fixed factors, as well as a Group by Session interaction. The pre-test and Control group were set as the reference level and the random effects structure included Subject, Item, and Place of Articulation (POA) of

the word-initial consonant as random intercepts[8]. Four models in total were implemented (one model per formant per vowel).

In the result sections that follow, the main effects of Group at the level of pre-test will be reported first in order to demonstrate the baseline standing of the three groups in terms of vowel shadowing. Thereafter, the interaction between Group and Session variables will be reported in order to show the differences between groups in the dynamics of their change in vowel production from pre- to post-test session. T-values were used to determine the significance of factors for all LME models using a significance criterion of $|t| > 2.00$. If the main model revealed significant effects/interactions, within group analysis and comparisons between experimental groups were completed post-hoc using pairwise comparisons with Tukey adjustments.

4.1.5 Data processing and analysis of VOT

In order to analyze voicing of target items, VOT was annotated from the release of the stop to the onset of voicing. Duration measurements were extracted using a custom Praat script, and all data was examined for outliers and corrected by hand, when necessary, prior to statistical analysis.

VOT of talkers' pre- and post-test shadowings were analyzed in three separate linear mixed effects models: one model for all voiceless VOT, one model for positive voiced VOT durations, and a model for negative voiced VOT durations. Positive and negative VOT durations of voiced stops were analyzed separately since combining the two would provide non-informative values that are not representative of actual speech (Lisker & Abramson, 1964). As voiceless VOT consists of only positive durations, it could be examined in one model. All mixed effects modeling was carried out using the lme4 package in R. Duration was submitted to each model as the dependent variable and Group (Audio-Only, Subtitled, and Control) and Session (pre-test and post-test) were used as fixed factors. The random effects structure included random intercepts of Talker and Item, and the Control group (pre-test) was the intercept for all models. Within group analysis and comparisons between experimental groups were completed post-hoc using pairwise comparisons

[8] In establishing the random effects structure, a maximal approach (Barr et al., 2013) was initially utilized that included random intercepts for Talker, Item, and POA, as well as random slopes for Group and Session. As the addition of random slopes caused singularity issues/convergence errors, the structure was simplified until the model converged. The resulting model was the maximum model that was able to be run without convergence issues. An identical procedure was followed throughout each analysis in the present chapter.

with Tukey adjustments. Significance for mixed effects models was established using t-values with a significance criterion of |t|> 2.00.

A frequency analysis of appropriate voicing categories was also conducted. Items were divided based on what category they fell into: prevoiced (durations less than 0 ms), short lag (durations greater than 0 ms, but less than 30 ms), and long lag (durations greater than 30 ms). Voiced and voiceless realizations were analyzed separately using two mixed effects binomial logistic regressions in order to establish whether category frequency significantly differed across groups and sessions. For the voiceless model, Category (corresponding to short lag and long lag realizations of /p/) was submitted as the binary dependent variable. The binary dependent variable, Category, in the voiced model consisted of prevoiced and short lag realizations of /b/. For both models (voiced and voiceless), Group (Audio-Only, Subtitled, and Control) and Session (pre-test and post-test) were submitted as fixed effects, and Subject and Item were included in the models' random effects structure. The reference category for both models was the Control group and pre-test. If the main model revealed significant effects/interactions, within group analysis and comparisons between experimental groups were completed post-hoc using pairwise comparisons with Tukey adjustments.

4.2 Results

4.2.1 Rounded vowels

A total of 5,328 items (74 talkers x 12 items x 3 repetitions x 2 sessions) containing rounded vowel /y/ or /u/ were collected from talkers across groups and sessions. During the data annotation and processing phase, 20 items (0.3%) were removed because of recording errors (poor recording quality, background noise, or additional interruptions prevented spectral analysis) and five (0.09%) were removed because they were outliers (three standard deviations away from the mean). This resulted in a total of 5,503 observations submitted to statistical analysis (2,654 items containing /y/ and 2,649 items containing /u/). The distribution of these observations across Vowel, Group (Audio-Only, Subtitled, and Control), and Session (pre-test and post-test) is presented in Table 4.4.

Table 4.4. Distribution of observations across Vowel (/y/ or /u/), Talker Group (Audio-Only, Subtitled, and Control), and Session (pre-test and post-test).

		/y/	/u/
Control	*Pre-test*	270	270
	Post-test	270	270
Audio-Only	*Pre-test*	532	533
	Post-test	539	539
Subtitled	*Pre-test*	521	518
	Post-test	522	519

Model talker acoustics

The model talker's vowel production was first examined to verify that it was representative of native French vowels. The model talker's vowel acoustics was in line with reports in previous literature that analyzed native vowel production (namely, Levy & Law, 2010; Strange et al., 2009). There was no overlap in terms of F2 between /y/ or /u/; /y/ was produced with a mean F2 of 12.14 Barks (SD= 0.26 Barks) and /u/ was produced with a mean F2 of 6.97 Barks (SD= 0.83 Barks), indicating a front and a back vowel, respectively. F1 for both /y/ and /u/ was similar and indicated a high vowel: 2.62 Barks (SD= 0.12 Barks) and 2.61 Barks (SD= 0.10 Barks), respectively.

Acoustic analysis of /y/

In the pre-test, as expected, all groups produced /y/ and /u/ as high vowels, as indicated by low F1 values with overlapping ranges. When averaged across all groups, at the pre-test, /y/ was produced with a mean F1 of 3.74 Barks (SD= 0.64; min: 1.71, max: 6.04), which is higher than that produced by the model talker. Therefore, if an improvement is to be observed as a result of film exposure, it is expected that F1 values will lower in the post-test for one or both of the experimental groups.

With regard to F1 values for /y/, results of a linear mixed effects model comparing the intercept (Control group) to the two experimental groups at the pre-test (Table 4.5) indicated that the Subtitled group produced significantly lower F1 values than the Control (t= -2.67). There was no difference between the Control group and the Audio-Only group prior to film exposure (t= -

1.12). These results suggest that despite the random assignment of participants to groups, and despite their lack of experience with French language, the Subtitled group was more target-like in their F1 values for /y/ at pre-test, compared to the Control group.

Following film exposure, participants' in both experimental groups lowered their F1 values for /y/ (Audio-Only: mean= 3.71 Barks, SD= 0.63 Barks; Subtitled: mean= 3.45 Barks, SD=0.68 Barks), while the Control group raised theirs (mean= 4.06 Barks, SD= 0.63 Barks) (Figure 4.1). Results of an LME model for /y/ revealed a significant Group by Session interaction (Control vs Audio-Only: t= -4.09; Control vs Subtitled: t= -4.15). The difference in the direction of change between the control and the two experimental groups suggests that the experimental groups achieved greater levels of convergence with the model talker and therefore produced /y/ in a more target-like manner following film exposure. Post-hoc pairwise comparisons revealed that all within group session effects were significant (Control pre-test vs. post-test: $p < .05$; Audio-Only pre-test vs post-test: $p < .005$; Subtitled pre-test vs post-test: $p < .001$). Finally, between group pairwise comparisons confirmed that film exposure had a positive effect on F1 production for /y/ regardless of access to subtitling (Audio-Only post-test vs Subtitled post-test: $p = .40$). Based on these results, it is concluded that film exposure had a positive effect on F1 (vowel height) for /y/, such that /y/ produced by the two experimental groups post-exposure became higher, as indicated by lower F1 values, thus approaching the vowel produced by the model talker.

Table 4.5. Table of fixed effects for the mixed effects model for /y/ F1 values (Barks).

Predictors	**Estimates**	**Std. Error**	**95% CI**	**Statistic (t-value)**
(Intercept)	4.00	0.14	[3.73, 4.27]	**28.86**
Audio-Only	-0.18	0.16	[-0.50, 0.14]	-1.12
Subtitled	-0.44	0.16	[-0.76, -0.12]	**-2.67**
Post-test	0.07	0.03	[0.01, 0.14]	**2.23**
Audio-Only: Post-test	-0.17	0.04	[-0.25, -0.09]	**-4.09**
Subtitled: Post-test	-0.17	0.04	[-0.25, -0.09]	**-4.15**

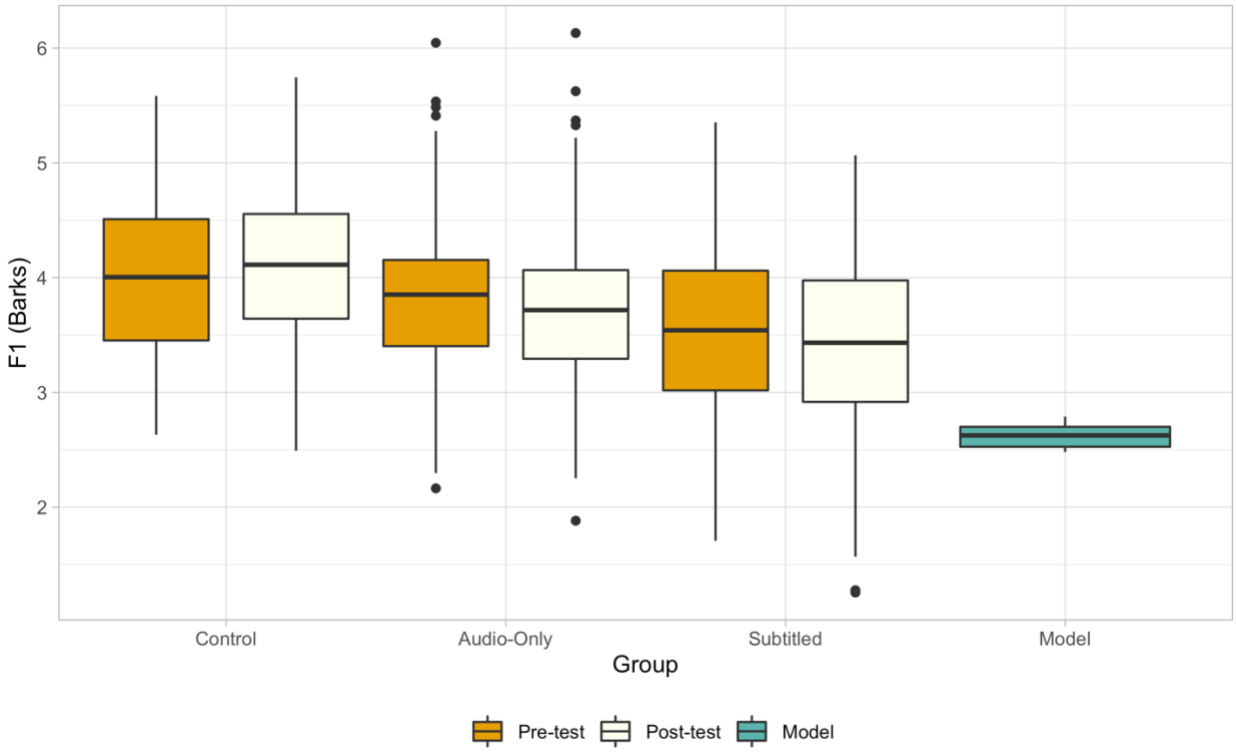

Figure 4.1. Bark values for /y/ across all groups (Control, Audio-Only, and Subtitled) and sessions (pre- and post-test) with the model talker for reference.

With regard to F2, prior to receiving film exposure, the participants did not produce overlapping ranges for /y/ and /u/ in terms of F2 values, realizing /y/ with a higher mean F2 of 12.16 Barks (SD= 1.29 Barks) and /u/ with a lower mean F2 of 9.33 Barks (SD= 1.49 Barks) (averaged across all groups). Assuming a 1500-1700 Hz (11-12 Barks) boundary distinguishing front and back vowels (Stevens, 2002), the participants clearly produce /y/ and /u/ as front and back, respectively, despite having no prior experience with French (results of participant shadowing of /u/ are outlined in greater detail in the next section). Additionally, French vowel /y/ was produced with F2 values similar to those of the model talker already in the pre-test (Figure 4.2)

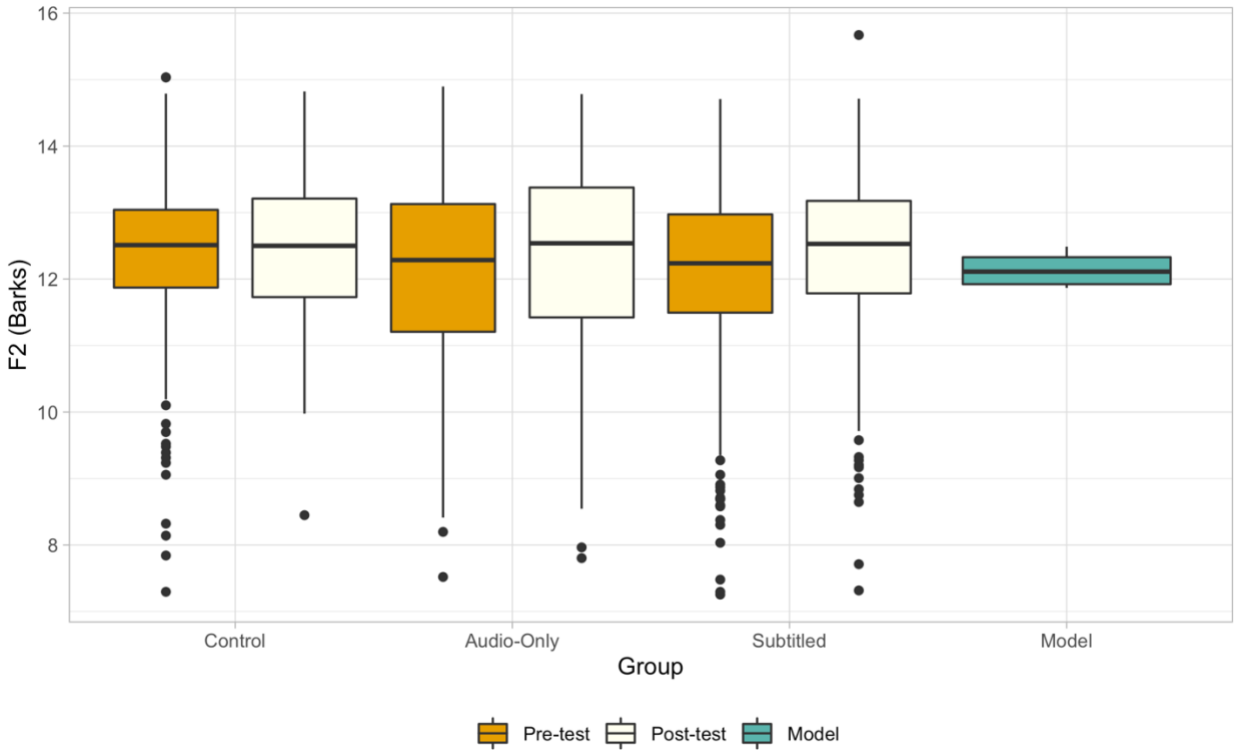

Figure 4.2. F2 Bark values for /y/ across all groups (Control, Audio-Only, and Subtitled) and sessions (pre- and post-test) with the model talker for reference.

As presented in Table 4.6, an analysis of the main effect of Group at the pre-test level revealed no differences between the Control group and the two experimental groups in terms of F2 (/y/: Control vs Audio-Only: t= -1.02; Control vs Subtitled: t= -0.85). These results suggest that all groups performed comparably before film exposure.

The participants also appear to introduce little change to their vowel production in terms of F2 following film exposure for /y/ (mean= 12.37 Barks, SD= 1.21 Barks). Results of the LME model for F2 for /y/ revealed no significant Group by Session interaction, indicating that exposure to film did not affect F2 values for experimental groups any differently than the Control group (Control vs Audio-Only: t= 1.03; Control vs Subtitled: t= 1.70).

Table 4.6. Table of fixed effects for the mixed effects model for /y/ F2 values (Barks).

Predictors	**Estimates**	**Std. Error**	**95% CI**	**Statistic (t-value)**
(Intercept)	12.23	0.33	[11.59, 12.87]	**37.36**
Audio-Only	-0.25	0.24	[-0.72, 0.23]	-1.02
Subtitled	-0.21	0.24	[-0.68, 0.27]	-0.85
Post-test	0.12	0.08	[-0.05, 0.27]	1.32
Audio-Only: Post-test	0.10	0.10	[-0.09, 0.30]	1.03
Subtitled: Post-test	0.17	0.10	[-0.03, 0.37]	1.70

Acoustic analysis of /u/

Turning to /u/, like /y/ participants produced the vowel with a mean of 3.68 Barks (SD=0.64; min: 1.20, max: 5.67) in the pre-test. Once again, as this average is higher than the model talker, improvement via film exposure would be demonstrated by lower F1 values for /u/ in the post-test.

Results of a linear mixed effects model that examined F1 values for /u/ (Table 4.7), indicated that, once again, the Subtitled group was significantly more target-like when compared to the Control group in the pre-test (t= -2.85), while the Audio-Only and Control groups performed similarly (t= -1.54).

Following film exposure, both experimental groups lowered their F1 values for /u/ in the post-test (Figure 4.3) (Audio-Only: mean= 3.69 Barks, SD= 0.60 Barks; Subtitled: mean= 3.51 Barks, 0.67 Barks), while the Control group raised theirs (mean= 4.02 Barks, SD= 0.62 Barks). In this case, however, results from the LME model demonstrate a significant interaction only when the Audio-Only group was compared to the Control group (t= -2.43). The Subtitled group did not differ significantly from the Control group in the post-test (t= -1.40). Therefore, improvement following film exposure for /u/ was confirmed for the Audio-Only group exclusively. This group demonstrates more successful convergence to the model talker in the post-test, by producing /u/ as

a higher vowel (as evidenced by lower F1 values). Pairwise comparisons did not reveal any additional significant within-group session effects, nor did it find the two experimental groups to perform differently in the post-test (p= .77).

Table 4.7. Table of fixed effects for the mixed effects model for /u/ F1 values (Barks).

Predictors	**Estimates**	**Std. Error**	**95% CI**	**Statistic (t-value)**
(Intercept)	3.96	0.14	[3.68, 4.24]	**27.66**
Audio-Only	-0.24	0.16	[-0.55, 0.07]	-1.51
Subtitled	-0.46	0.16	[-0.77, -0.14]	**-2.85**
Post-test	0.05	0.03	[0.01, 0.11]	1.54
Audio-Only: Post-test	-0.10	0.04	[-0.18, -0.02]	**-2.43**
Subtitled: Post-test	-0.06	0.04	[-0.14, -0.02]	-1.40

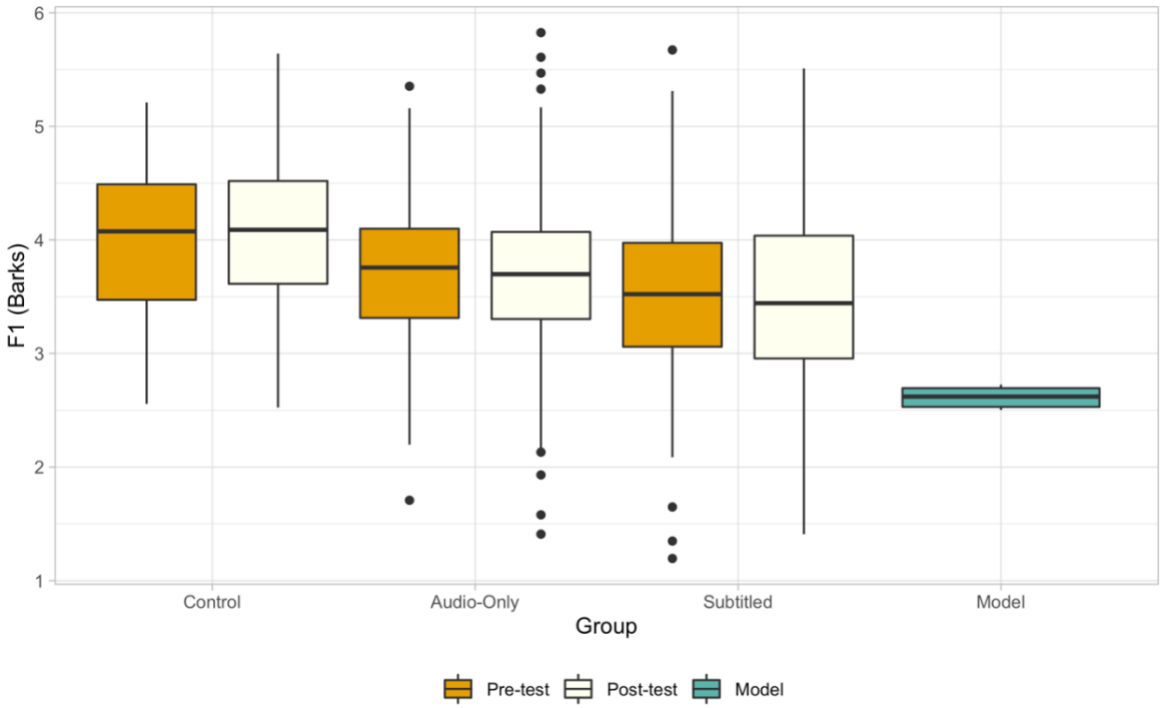

Figure 4.3. F1 Bark values for /u/ across all groups (Control, Audio-Only, and Subtitled) and sessions (pre- and post-test) with the model talker for reference.

As mentioned in the previous section, F2 values for /u/ revealed that participants produced it as a back vowel in the pre-test, however it was produced as a fronter vowel by English-speaking participants when compared to the native French model (pre-test mean: 9.33 Barks, SD= 1.49 Barks). Therefore, while little space for post-test improvement remained for /y/ due to its initial closeness to the model, film exposure for /u/ would be expected to result in a more retracted vowel (lower F2), if an improvement is observed.

Like F2 values of /y/, an analysis of the main effect of Group at the pre-test level did not reveal any differences between the three groups with respect to vowel backness (F2) Control vs Audio-Only: t= -0.63; Control vs Subtitled: t= -1.46) (Figure 4.4). These results suggest that all groups performed comparably before film exposure.

As presented in Table 4.8, despite there being room for improvement in F2 shadowing of /u/ following film exposure, results of the LME model for F2 of /u/ did not indicate a significant difference between experimental groups vs. the Control group following film exposure (Control vs Audio-Only: t= -0.52; Control vs Subtitled: t= -0.08).

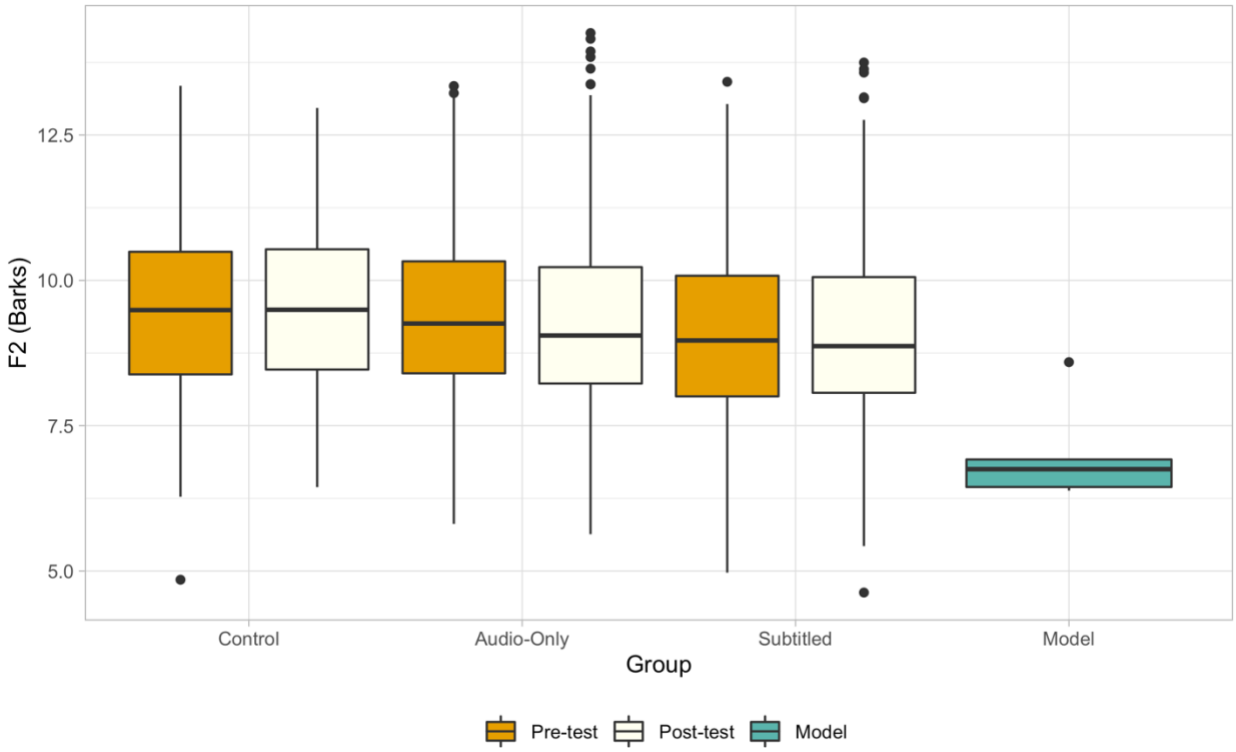

Figure 4.4. F2 Bark values for /u/ across all groups (Control, Audio-Only, and Subtitled) and sessions (pre- and post-test) with the model talker for reference.

Table 4.8. Table of fixed effects for the mixed effects model for /u/ F2 values (Barks).

Predictors	**Estimates**	**Std. Error**	**95% CI**	**Statistic (t-value)**
(Intercept)	9.79	0.63	[8.54, 11.03]	**15.40**
Audio-Only	-0.17	0.27	[-0.70, 0.36]	-0.63
Subtitled	-0.40	0.27	[-0.93, 0.14]	-1.46
Post-test	0.03	0.09	[-0.20, 0.14]	-0.33
Audio-Only: Post-test	-0.06	0.11	[-0.27, 0.16]	-0.52
Subtitled: Post-test	0.01	0.11	[-0.21, 0.22]	0.08

Summary of results for rounded vowels

Given the pre-test positioning of participants' vowels in the vertical dimension with respect to the model talker's production, lower F1 values were predicted post-exposure, resulting in higher vowels, if an improvement was to be observed. This prediction was upheld. Both experimental groups produced a higher /y/ (with lower F1 values) in the post-test and the Audio-Only group produced a higher /u/ (lower F1 values) in the post-test, when compared to the direction and rate of change for the Control group.

An improvement for /y/ in terms of backness, was not anticipated given that this vowel was already produced with F2 values comparable to those of the native French model in the pre-test. This assumption was confirmed. Conversely, for /u/, a more retracted vowel (lower F2) produced by the experimental groups following exposure was expected, given that participants' pre-test realizations of French /u/ were more fronted than the model. This prediction, however, was not borne out. Participants in the experimental groups failed to implement a post-exposure change in the realization of /u/ in terms of backness.

Thus, two main conclusions can be derived. First, the production of /y/ was more target-like in the pre-test than the production of /u/. Second, a post-exposure improvement for /y/ was observed whenever one would be expected, but /u/ only exhibited post-exposure improvement for F1/height (no improvement in F2/backness) and only in one of the experimental groups, Audio-Only.

4.2.2 VOT

Across groups and sessions, 5,328 (74 talkers x 12 items x 3 repetitions x 2 sessions) observations were collected. Approximately 5.18% (276 total observations) of the collected items were excluded from the final statistical analyses primarily due to severe mispronunciations (e.g. talker did not pronounce the word initial consonant, produced a different consonant, etc.). With respect to the distribution of observations submitted to each statistical model, 2,563 items were produced with voiceless VOT (i.e. /p/) and 2,489 items were produced with voiced VOT (i.e. /b/). Of the voiced items, 1,000 were realized with positive durations (i.e. short lag) and the remaining 1,489 were produced with negative durations (i.e. prevoicing). Table 4.9 presents the distributions of the

items submitted to the statistical models with respect to Voicing (voiced and voiceless), Group (Audio-Only, Subtitled, and Control), and Session (pre-test and post-test).

Table 4.9. Distribution of observations across Voicing (voiced and voiceless), Talker Group (Audio-Only, Subtitled, and Control), and Session (pre-test and post-test).

		Voiced		**Voiceless**
		Positive	*Negative*	*Positive*
Control	*Pre-test*	132	125	266
	Post-test	134	118	268
Audio-Only	*Pre-test*	174	312	503
	Post-test	204	305	513
Subtitled	*Pre-test*	194	296	507
	Post-test	162	333	506

Model talker acoustics

Preceding the acoustic analysis, model talker VOT productions were examined to confirm that they were representative of native French norms. The model talker realized all voiceless stops with short lag VOT with an mean duration of 21.34 ms (SD= 8.54 ms). All voiced stops were produced with prevoicing with an average duration of -170.18 ms (SD= 46.83 ms). With this in mind, the model talker appeared to produce native-like VOT in line with previously established findings (Caramazza & Yeni-Komshian, 1974).

Voice onset time duration for voiceless stops

In order to determine if foreign film exposure had an effect on voiceless VOT duration, all voiceless items were submitted to a linear mixed effects model with main effects of Group (Audio-Only, Subtitled, and Control), Session (pre-test and post-test), and Group by Session interaction (Table 4.10). With regard to pre-test performance, the statistical analysis did not determine any significant differences between groups (Audio-Only: t= -0.58; Subtitled: t= -1.68), indicating that

all groups produced voiceless stops with relatively equal durations prior to film exposure (Audio-Only: mean= 53.66 ms, SD=35.62 ms; Subtitled: mean= 47.53 ms, SD= 29.27 ms; Control: mean= 56.25 ms, SD= 35.63 ms). These values can be characterized as long lag VOT and if an improvement in the direction of approximating the model is observed post-exposure, these VOT values are expected to shorten.

Results of the LME model additionally show that there was not a significant effect of Session (Post-test: t= -1.79), suggesting that despite the apparent shortening of voiceless VOT durations across all groups in the post-test (Audio-Only: mean= 48.11 ms, SD= 34.92 ms; Subtitled: mean= 44.64 ms, SD= 29.41 ms; Control: mean= 52.63 ms, SD= 33.34 ms) (Figure 4.5), this was a marginal practice effect. There were also no significant interactions in the model, indicating that the two experimental groups did not outperform the Control group following exposure to foreign film. Together these findings suggest that film exposure did not cause talkers to produce significantly more target-like voiceless VOT.

Table 4.10. Table of fixed effects for the mixed effects model of voiceless VOT duration.

Predictors	**Estimates**	**Std. Error**	**95% CI**	**Statistic (t-value)**
(Intercept)	55.99	8.81	[38.72, 73.26]	**6.35**
Audio-Only	-2.96	5.09	[-12.94, 7.01]	-0.58
Subtitled	-8.59	5.11	[-18.61, 1.43]	-1.68
Post-test	-3.54	1.99	[-7.41, 0.33]	-1.79
Audio-Only: Post-test	-1.58	2.44	[-6.36, 3.20]	-0.65
Subtitled: Post-test	0.10	2.44	[-4.61, 4.96]	0.07

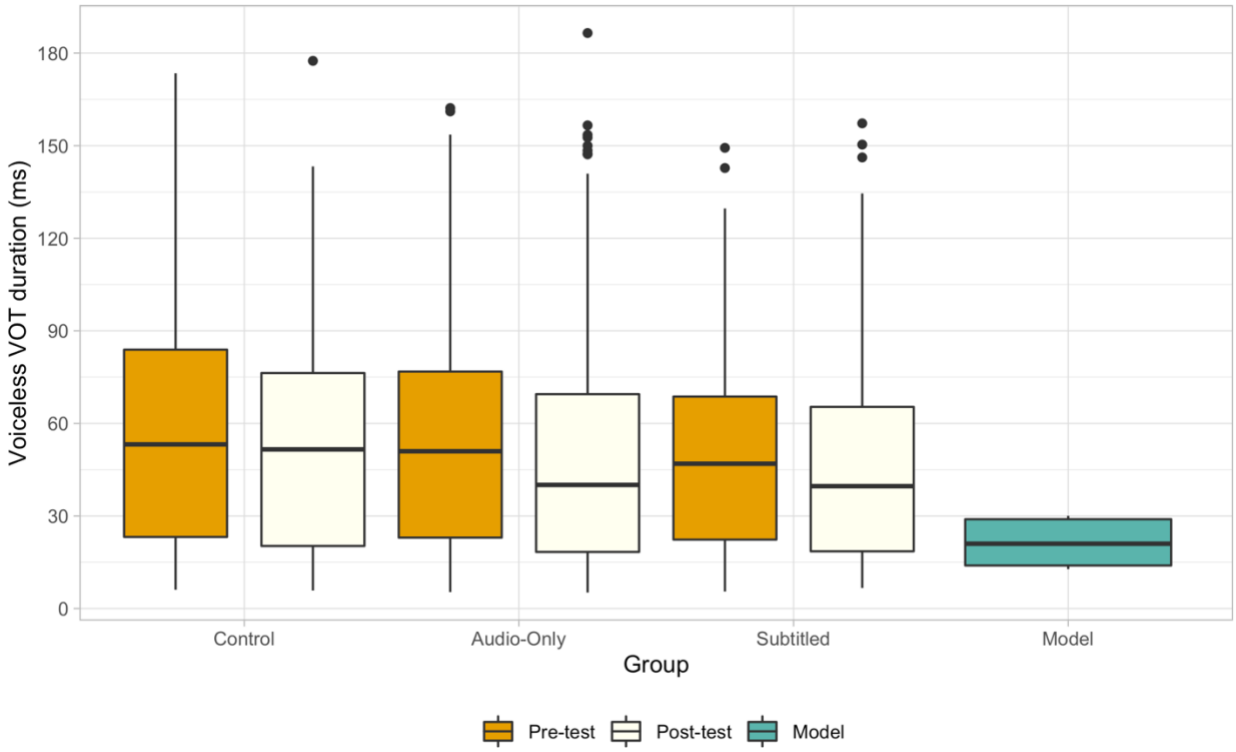

Figure 4.5. Voiceless VOT duration across groups and sessions with the model talker for reference.

Voiceless category frequency

In addition to the duration analysis, a supplementary analysis was completed in order to determine if talkers produced more instances of the target French voiceless category, short lag, following film exposure. If exposure to French through film had an effect on voiceless category development, a greater number of tokens classified as short lag (VOT duration less than 30 ms) would be expected. A mixed effects binomial logistic regression (Table 4.11) did not reveal any significant differences between groups in the post-test, nor did it indicate a significant main effect of Session. Additionally, despite a visible increase in short lag frequency by both experimental groups (Figure 4.6), neither Group*Session interaction was significant. These findings demonstrate that while marginal attempts to realize more French voiceless stops as short lag in the post-test were made, especially in the Audio-Only group, practice and film exposure did not significantly affect voiceless category development.

Table 4.11. Table of fixed effects for the mixed effects binomial logistic regression of voiceless VOT categories.

Predictors	**Odds Ratios**	**Std. Error**	**95% CI**	**Statistic (z-value)**	***p*-value**
Intercept	5.09	0.88	[0.90, 28.74]	1.84	0.07
Audio-Only	0.66	0.54	[0.23, 1.89]	-0.77	0.44
Subtitled	0.62	0.54	[0.22, 1.79]	-0.88	0.38
Post-test	0.72	0.26	[-0.43, 1.20]	-1.27	0.20
Post-test: Audio-Only	0.75	0.32	[0.40, 1.39]	-0.92	0.36
Post-test: Subtitled	0.89	0.32	[0.48, 1.66]	-0.36	0.72

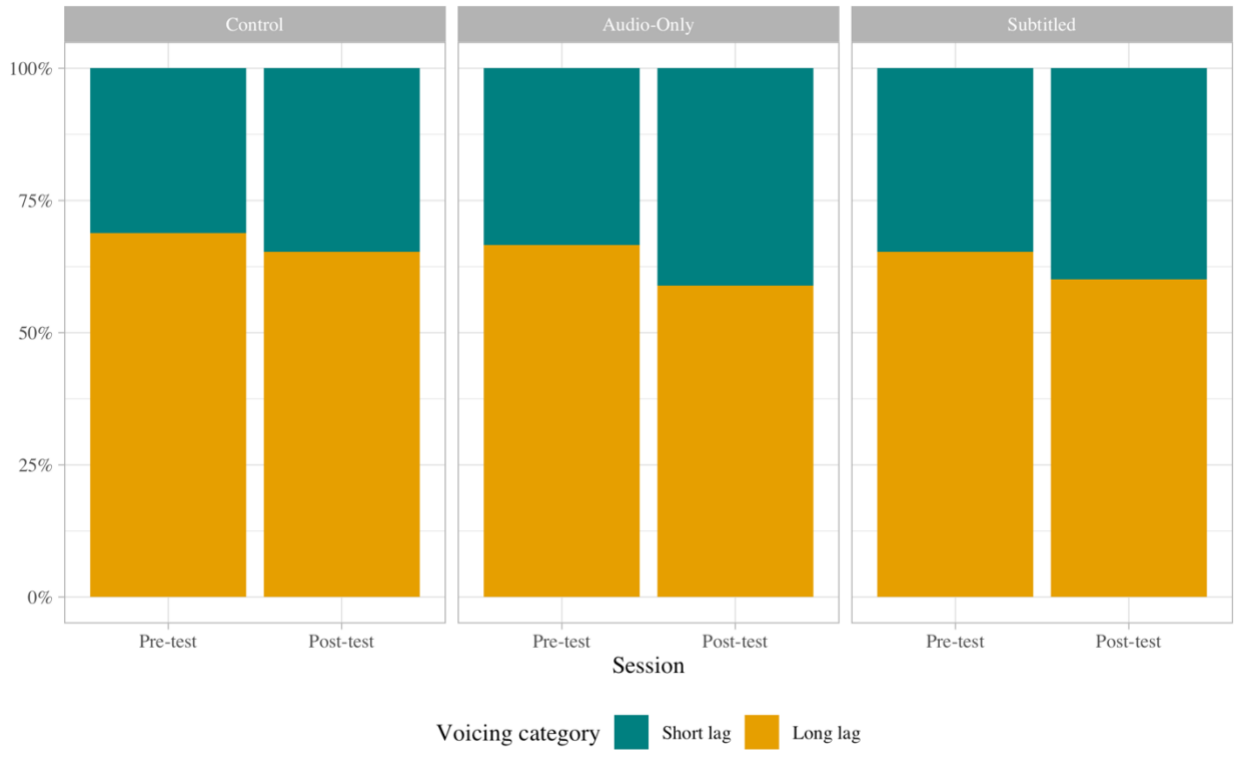

Figure 4.6. Frequency of voiceless categories (short and long lag) across groups and sessions.

Voice onset time duration for voiced stops

To examine if film exposure affected VOT duration of voiced stop, /b/, two linear mixed models were run: one for positive voiced durations (i.e. short lag) and a separate one for negative voiced durations (i.e. prevoicing). As French realizes voiced VOT as negative, a positive effect of film exposure would be represented by significantly shortened positive voiced durations post-exposure (representative of an attempt to shorten VOT in order to potentially become negative). It is unlikely that negative voiced durations would significantly change due to exposure, but this effect may be present if there was a larger number of stops categorized as prevoiced in the post-test (more prevoiced tokens could lead to a significant difference in duration due to larger amounts of variation). Both models had the same structure with Group and Session as fixed effects and interactions, as well as Subject and Item random intercepts. As demonstrated in Table 4.12, with regard to positive voiced VOT, all groups performed similarly to one another in the pre-test (Audio-Only: t= 0.81; Subtitled: t= 0.93) (Audio-Only: mean= 17.86 ms, SD= 5.70 ms; Subtitled: mean= 17.61 ms, SD= 5.18 ms; Control: mean= 16.60 ms, SD= 5.77 ms). In addition, there was no significant main effect of Session, suggesting that regardless of group assignment, there was no significant change in duration between sessions (Audio-Only: mean= 17.26 ms, SD= 6.16 ms; Subtitled: mean= 17.23 ms, SD= 5.86 ms; Control: mean= 17.23 ms, SD= 5.06 ms). Furthermore, despite visible shortening of average positive voiced durations by both experimental groups (Figure 4.7) in the post-test, this trend is not significant when compared to the Control group (Audio-Only vs. Control: t= -1.47; Subtitled vs. Control: t= -0.99). With these results in mind, there is no present evidence to suggest any potential effect of film exposure on voiced stops produced with positive VOT durations.

Table 4.12. Table of fixed effects for the mixed effects model of positive voiced VOT duration.

Predictors	Estimates	Std. Error	95% CI	Statistic (t-value)
(Intercept)	17.40	1.27	[14.91, 19.88]	**13.72**
Audio-Only	0.85	1.05	[-1.22, 2.92]	0.81
Subtitled	0.99	1.06	[-1.09, 3.06]	0.35
Post-test	0.26	0.57	[-0.87, 1.38]	0.65
Audio-Only: Post-test	-1.12	0.76	[-2.61, 0.37]	-1.47
Subtitled: Post-test	-0.76	0.77	[-2.26, 0.74]	-0.99

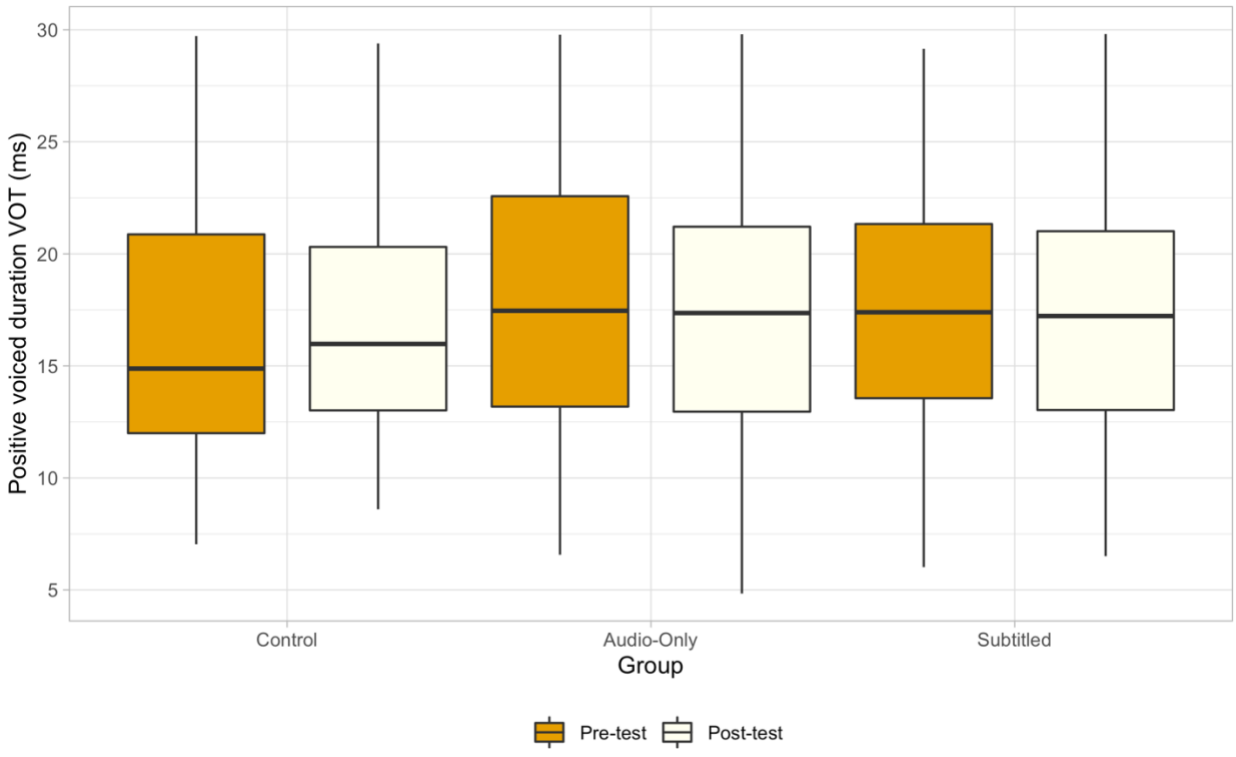

Figure 4.7. Positive voiced VOT duration across groups and sessions.

Regarding voiced stops produced with negative VOT (LME results presented in Table 4.13), all groups appear to perform comparably in the pre-test (Audio-Only: mean= -151.05 ms, SD= 55.49 ms; Subtitled: mean= -141.75 ms, SD= 59.08 ms; Control: mean= -159.12 ms, SD= 51.27 ms), as evidenced by insignificant main Group effects (Audio-Only: t= -0.01.; Subtitled: t= 0.31). A significant main effect of Session revealed that all groups (t= -3.31), regardless of film exposure, significantly lengthened their negative voiced VOT durations in the post-test, indicative of a practice effect (Audio-Only: mean= -156.08 ms, SD= 58.11 ms; Subtitled: mean= -150.42 ms, SD= 66.28 ms; Control: mean= -176.10 ms, SD= 58.21 ms) (Figure 4.8). The present interaction between Session and Group (Audio-Only) illustrates that film exposure caused the Audio-Only group to produce significantly different VOT durations when compared to the Control group (t= 2.07). In this case, while the Control group greatly lengthened their prevoiced durations in the post-test, the Audio-Only group only modestly lengthened theirs. A similar trend is found in the interaction between Session and Group for the Subtitled group, though the effect is only marginal (t= 1.92). Pairwise comparisons did not reveal any additional findings, as all within and between group comparisons were not significant.

Table 4.13. Table of fixed effects for the mixed effects model of negative voiced VOT duration.

Predictors	**Estimates**	**Std. Error**	**95% CI**	**Statistic (t-value)**
(Intercept)	-146.88	12.45	[-171.29, -122.47]	**-11.79**
Audio-Only	-0.07	12.52	[24.61, 28.63]	-0.01
Subtitled	3.87	12.63	[-20.88, 28.63]	0.31
Post-test	-21.13	6.39	[-33.66, -8.61]	**-3.31**
Audio-Only: Post-test	15.64	7.56	[0.81, 30.46]	**2.07**
Subtitled: Post-test	14.44	7.53	[-0.34, 29.22]	1.92

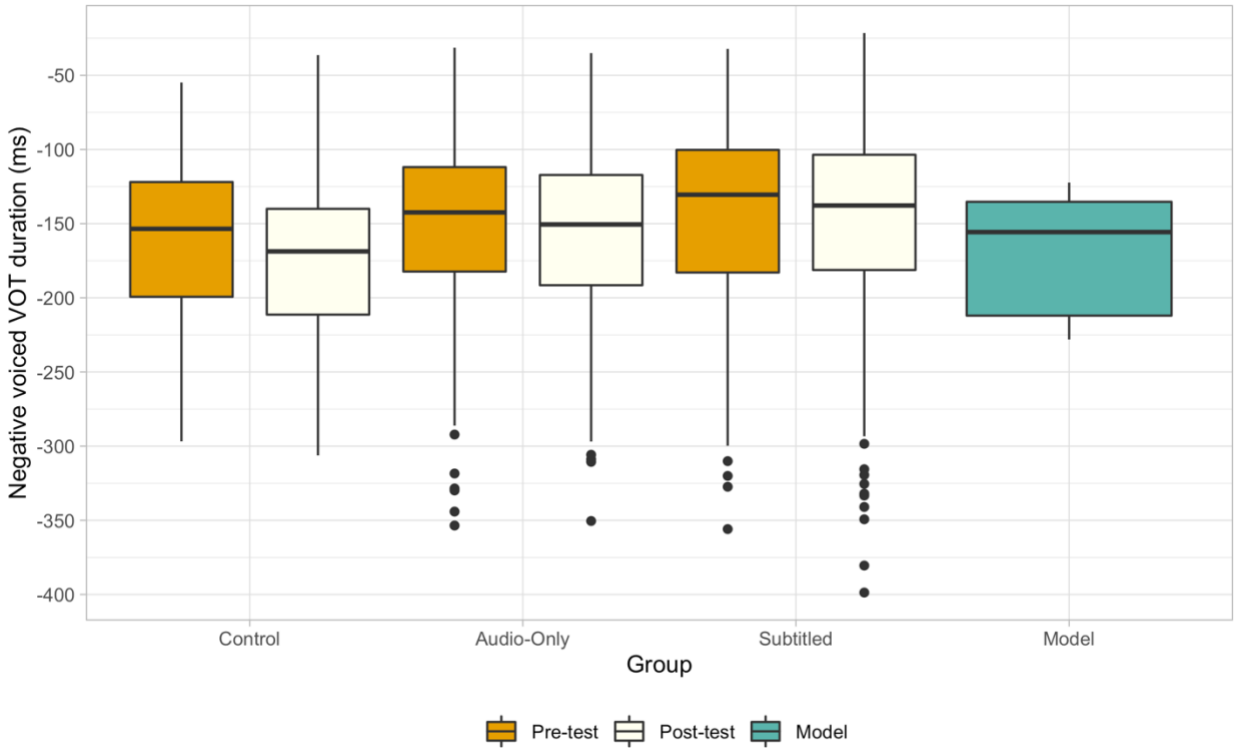

Figure 4.8. Negative voiced VOT duration across groups and sessions.

Voiced category frequency

In order to determine if exposure to film had an effect on frequency of voiced VOT categories (prevoiced vs. short lag), a mixed effects binomial logistic regression was completed. If film exposure had a positive effect on voiced category frequency, a larger number of target-like prevoiced stops would be expected. Initially, as evidenced in Table 4.14, results examining pre-test productions revealed while the Subtitled group performed similarly to the Control group (p= .09), the Audio-Only group realized more French /b/s as prevoiced than the Control group (p< .05). A main effect of Session did not appear to be significant in the model (p= .59), nor were there any Group*Session interactions (Audio-Only vs. Control: p= .82; Subtitled vs. Control: p= .07). With these results in mind, despite Figure 4.9 demonstrating the Subtitled group's increase in French /b/s produced with target-like prevoicing in the post-test there was no statistical evidence to support the claim that film exposure was effective at developing target-like voicing categories.

Table 4.14. Table of fixed effects for the mixed effects binomial logistic regression of voiced VOT categories.

Predictors	**Odds Ratios**	**Std. Error**	**95% CI**	**Statistic (z-value)**	***p*-value**
Intercept	1.39	0.49	[0.53, 3.66]	0.67	0.50
Audio-Only	0.35	0.54	[0.12, 1.00]	-1.96	**0.05**
Subtitled	0.40	0.54	[0.14, 1.17]	-1.67	0.09
Post-test	1.14	0.25	[0.70, 1.85]	-0.54	0.59
Post-test: Audio-Only	1.07	0.29	[0.61, 1.87]	-0.23	0.82
Post-test: Subtitled	0.59	0.29	[0.33 – 1.05]	-1.78	0.07

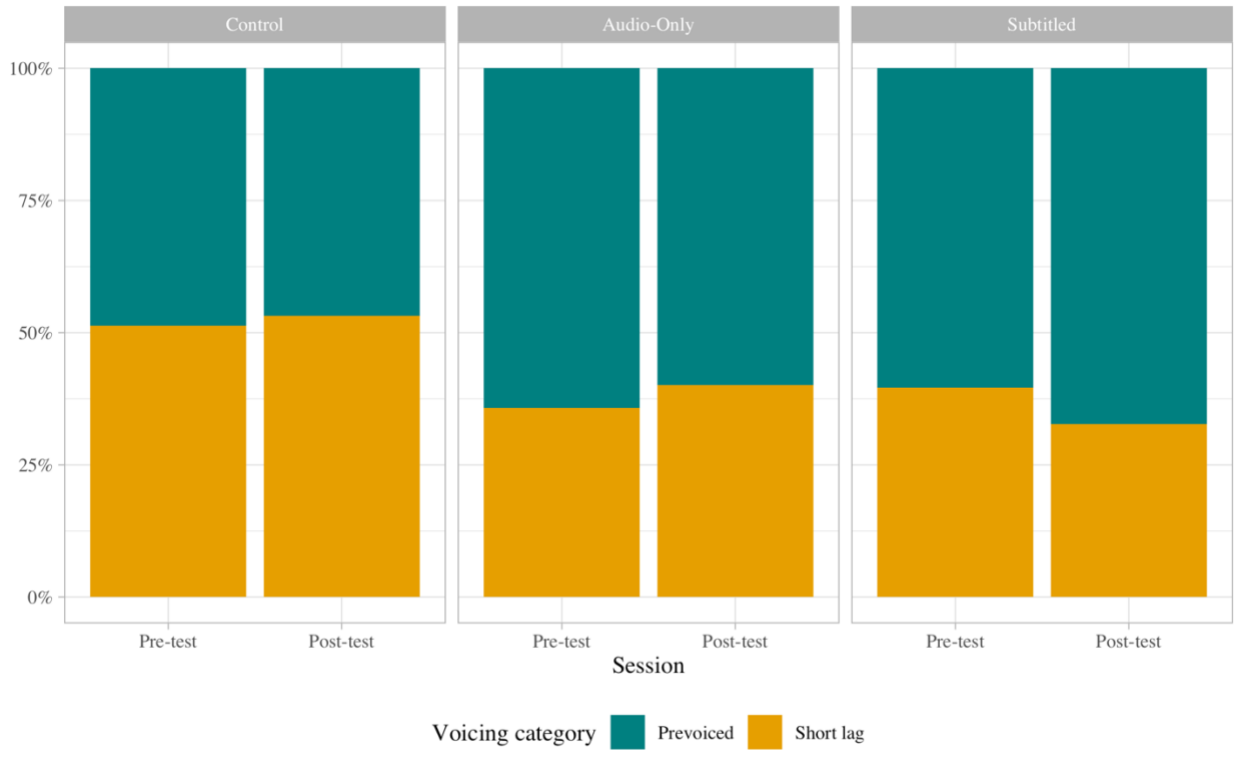

Figure 4.9. Frequency of voiced categories (prevoiced and short lag) across groups and sessions.

Summary of results for VOT

While the findings in the current experiment suggest minimal improvements in non-native VOT shadowing in the post-test, as a whole, VOT appears to be relatively unaffected by practice and film exposure. That said, despite not reaching significance, the two experimental groups do appear to trend in the expected direction in the post-test, and it is possible that further exposure to foreign film would cause these marginal improvements to become significant. For example, the Audio-Only group produces shorter voiceless VOT and more instances of target-like short lag in the post-test. The Subtitled group also follows similar trends (shorter voiceless duration and more frequent short lag). Regarding positive voiced durations, both experimental groups appear to make marginal attempts at shortening their VOT in the post-test, while the Control group lengthens theirs. Though this finding is not significant, it suggests that when French voiced VOT is erroneously produced with positive durations, talkers are making some attempt to shorten it. Negative voiced durations do appear to be affected by exposure to film, albeit in the direction opposite as predicted, as the Audio-Only group produces significantly shorter prevoicing than the Control group in the post-test. This result is nearly replicated by the Subtitled group ($t = 1.92$) but does not reach significance. In a frequency analysis of voiced categories, the Subtitled group trends towards more target-like voiced category realization (Control and Audio-Only produce less prevoicing in the post-test), but this result was not confirmed statistically. The aforementioned analysis additionally appears to be the only instance where the Subtitled group outperforms or demonstrates additional benefit over the Audio-Only group. That said, despite the marginal advantage the Audio-Only group displays throughout the analyses, these results are not statistically significant.

4.3 Interim Discussion

The current chapter examines the effect of exposure to foreign film on non-native speech production by specifically investigating how monolingual English participants shadow two French variables, high rounded vowels, /y/ and /u/, and voiced and voiceless bilabial VOT. Two research questions (RQ1 and RQ2) and hypotheses addressing the production component of the present study were posed in Chapter 2. A discussion of these questions/hypotheses with regard to the current results is presented in the sections that follow.

4.3.1 Effect of film on rounded vowel shadowing

With regard to high rounded vowel shadowing, the following research question and hypothesis was posed:

(RQ1) Will naïve, monolingual American English speakers shadow word medial French high rounded vowels (/y/ and /u/) in a more target-like manner after exposure to foreign film?

Hypothesis 1: Rounded vowels are hypothesized to be particularly susceptible to influence following film exposure as they provide the talker with additional, beneficial visual information (lip rounding) on top of an acoustic signal (Traunmüller & Öhrström, 2007). This hypothesis is also supported by literature that supports the idea that vowel production can be improved with an audiovisual training procedure (Inceoglu, 2016). While Inceoglu (2016) did not find a difference between nasal vowel production by those trained with an audio-only modality when compared to participants who recieved audio and visual, the additional visual cues offered by rounded vowels are likely to enhance production.

Additionally, previous research on the production of French /y/ and /u/ by native English talkers with varying amounts of French experience, demonstrates that "new" vowel /y/ is easier to produce than "similar" vowel /u/ (Flege, 1987; Flege & Hillenbrand, 1984; Levy & Law, 2008, 2010; Rochet, 1995). With that in mind, it is likely that talkers will improve more at /y/ shadowing following exposure to foreign film as it does not already exist in the English vowel inventory (Flege, 1995, 2002).

In general, an acoustic analysis revealed a significant positive effect of film exposure on both experimental groups' productions of /y/, and, to some degree, /u/ (RQ 1). This pattern demonstrates that watching a film in French, with or without subtitles, leads to greater convergence towards target-like French norms by non-native participants.

For /u/ specifically, film exposure affected participants in the Audio-Only group but not in the Subtitled group, and in one acoustic dimension (F1/height) only. Part of the explanation for this pattern possibly lies in the fact that the Subtitled group began the experiment with a significantly more French-like realization of both vowels in terms of height compared to other groups. This more advanced initial position could have prevented the Subtitled group from substantial improvement across sessions. However, a more important observation is that both

experimental groups failed to improve the pronunciation of /u/ in terms of backness, despite the lack of any pre-existing advantages when compared to the Control group. Thus, the lack of improvement for /u/ must be conditioned by deeper reasons. It should be noted that while French /y/ also failed to improve in backness post-exposure, /y/ was pronounced with essentially native-like F2 already in the pre-test, ruling out the possibility of improvement due to the ceiling effect.

A theoretically-based explanation is offered by the SLM (Flege, 1987, 2002, 1995). The model divides L2 speech categories into two types: those, for which acoustically and perceptually similar L1 categories can be found, and those for which L1 counterparts do not exist. French /u/ belongs to the first type, since English /u/ is acoustically and functionally comparable to French /u/. In contrast, French /y/ is a novel category for English-speaking learners. The SLM posits that native speakers of English will experience difficulty in detecting subtle acoustic differences between French /u/ and English /u/ and are therefore likely to equate the French vowel to the English one, also producing it as such. Conversely, acquisition of French /y/ is predicted to be easier for English-speakers since /y/ does not have a close counterpart in the L1.

In the current study, the existing L1 /u/ category appears to be preventing participants from producing F2 values for French /u/ in a more target-like manner, pre- or post-exposure, while participants' /y/ productions effectively converge with model's /y/ in terms of backness in both sessions. It is not clear, however, why naïve participants managed to improve their French /u/ post-exposure in terms of height but failed to improve it in terms of backness. Further research needs to be conducted in order to determine if longer and/or more frequent exposure to native speech via film can aid in neutralizing the interference effects of English /u/ on French /u/.

4.3.2 Effect of film on VOT shadowing

In addressing the effect of foreign film on VOT shadowing, the following research question and hypothesis was developed:

(RQ2) Will naïve monolingual American English speakers shadow French voicing of bilabial stops in word-initial position ([+voice] prevoiced, [-voice] short lag) in a more target-like manner after exposure to foreign film?

Hypothesis 2: In comparison to rounded vowels, improvement in VOT production is likely to be more challenging to talkers due to the lack of visual saliency (Inceoglu, 2016). That said, it is hypothesized that talkers will demonstrate some changes in VOT duration and category frequency following exposure to foreign film, though these changes may be small and talkers are likely to rely heavily on L1 transfer during both sessions (Flege & Eefting, 1988; Nielsen, 2011).

With regard to RQ2, there does not appear to be any evidence in the current analysis that exposure to foreign film has a significant effect on VOT shadowing, ultimately failing to provide support for Hypothesis 2.

Results from an analysis of voiceless stops indicate that exposure to foreign film has little effect on duration or voiceless category frequency, as demonstrated by insignificant interactions in their respective models. That said, following film exposure, both experimental groups visibly attempt to shorten their durations despite the change not reaching significance. It is possible that with further film exposure, the present difference will become larger, ultimately providing further evidence to support the use of film in L2 speech development.

Regarding voiced stops, the duration of positive voiced VOT does not appear to be affected by exposure to film. This is perhaps unsurprising, as French foreign film would not contain positive voiced VOT, so talkers would not have a guide for potential durational differences. When French voiced stops are erroneously produced with positive durations, both experimental groups visibly shorten their durations in the post-test and the Control group lengthens theirs. This might be the start of an attempt by the experimental group to reduce positive VOTs of voiced stops as much as possible even if prevoicing is not achieved, however as this finding was not statistically significant, further research would need to be completed to provide confirmation.

When voiced stops are realized with prevoicing, film does appear to significantly effect post-test durations, but only for the Audio-Only group. In this case, film creates a smaller increase in duration for exposed talkers than those who are not exposed (the Subtitled group also patterns similarly to the Audio-Only group, but the effect does not reach significance). In general, the explanation for this finding is not particularly clear. It is possible that film exposure causes the two experimental groups to clue into the fact that prevoicing duration itself does not particularly matter and that prevoicing frequency is more important. However, if this was the case, one would expect

experimental group prevoicing frequency to significantly increase following film exposure, which does not occur. In fact, the Audio-Only group appears to marginally decrease frequency of prevoicing from pre- to post-test. It is instead possible that this effect might be present because of a few Control group talkers who are strongly influenced by a practice effect and ultimately cause the Control group post-test average duration to increase. If this is the case, prevoicing duration is not necessarily influenced by film exposure, but is instead affected by individual talker variation.

4.3.3 Exposure to film and non-native speech shadowing

In general, results from analyses of pre-test rounded vowel and VOT shadowing, indicate that there is some variation in the groups' ability to converge with the model talker. This finding is particularly present in /y/ and /u/ F1 productions by the Subtitled Group and frequency of prevoicing by the Audio-Only group. Further exploration into extralinguistic factors that have previously been shown to affect shadowing capability would need to be completed in order to make conclusions as to why this difference was found.

With regard to the role foreign film exposure plays in non-native speech shadowing, the results are highly dependent on the variable examined. In the present study, French rounded vowels /y/ and /u/ appear to be significantly affected by exposure to film, while VOT only exhibits insignificant trends. It is possible that these contrasting findings are present because vowel quality is more perceptually salient to non-native talkers than a temporal measure like VOT (Inceoglu, 2016). This is supported by previous research that demonstrates the juxtaposition between the relative ease of vowel shadowing by naïve, monolinguals speakers (Levy & Law, 2010), and the challenge of non-native VOT shadowing (Flege & Eefting, 1988; Nielsen, 2011). In this case, perceptual saliency, along with the additional articulatory information that rounded vowels provide, likely cause a boost in development that is not available for VOT. Non-significant trends in the current VOT analyses however, suggest that VOT development via film exposure is not impossible, but that it might simply progress slower than non-native vowel production. Further research that explores greater film exposure should be completed in order to confirm this hypothesis.

CHAPTER 5. NATIVE LISTENER PERCEPTUAL JUDGEMENTS

Chapter 4 presented an acoustic analysis of the effect of foreign film exposure on non-native shadowing of French high rounded vowels, /y/ and /u/, and bilabial consonant voicing (VOT). While this analysis of acoustic attributes is paramount in establishing the precise nature of the effects of foreign film exposure on non-native speech shadowing, it provides an incomplete picture as it can potentially collapse the multi-dimensional nature of shadowing into a few acoustic parameters (Pardo, 2013). For example, in the current study, improvements in consonant production are measured by analyzing VOT duration and category frequency. While VOT is the primary correlate of voicing and is perhaps the most informative in this particular analysis, it is by no means the only acoustic parameter that talkers could adjust in an attempt to become more target-like. For example, secondary correlates like onset f0 (fundamental frequency at the start of voicing; Abramson & Lisker, 1965; Hombert, 1976; House & Fairbanks, 1953; Kingston & Diehl, 1994; Ohde, 1984), amplitude of aspiration noise (Repp, 1979), oral closure duration (Keating, 1984) and F1 cutback (F1 transition offset time; Liberman et al., 1958; Stevens & Klatt, 1974) can also be altered by speakers. Therefore, in order to gain a more comprehensive view of speech shadowing following exposure to foreign film, this study enhanced the shadowing experiment by collecting additional perceptual measures of convergence (i.e. native listener perceptual judgements). The ultimate goal of the current experiment is to determine if acoustic adjustments (F1, F2, VOT, or others) that non-native talkers made following exposure to foreign film were perceptible to native talkers.

The present chapter is divided into the following sections: Section 5.1 presents methods specific to the present experiment, including detailed information about the native listeners recruited to participate in the study (section 5.1.1), the materials used during the task (section 5.1.2), all task procedures (section 5.1.3), and analysis (section 5.1.4). Section 5.2 presents the results of the experiment and is divided based on judgements of trials containing rounded vowels (section 5.2.1) and trails containing word-initial bilabial stops (section 5.2.2). Finally, section 5.3 details relevant discussion points brought forward by results in the current experiment.

5.1 Methods

5.1.1 Listeners

222 listeners (135 male, 84 female, and 3 non-binary; mean age= 28.18 y.o., SD= 9.40) were recruited using Prolific (Prolific, Oxford, UK), an online data collection platform. Listeners were pre-screened based on their first language and only participants who indicated that French was their first language were recruited. While the majority of listeners were born in and/or currently resided in France, recruitment was not restricted only to France-based participants[9]. Country of birth and country of current residence across all listeners is presented in Table 5.1 and Table 5.2, respectively. As Prolific is a United Kingdom-based website, all listeners had some knowledge of English (mean age of acquisition 8.90, SD 3.36), with a majority of them being highly proficient. When asked to rate their proficiency in English using a sliding Likert scale (1-7, 1 being "Very Poor" and 7 begin "Native-like"), the average was 5.75 (SD= 1.14). 115 participants reported knowledge of additional languages other than French and English with Spanish and German being the most common languages. None of the participants reported history of a speech or hearing disorder. Each participant was compensated $1.40 total to complete the 12-minute task ($7.00/hr).

[9] While a homogenous group of native French speakers who were born and currently residing in a single country (i.e. France) would have been preferred in order to control for potential confounding variables, Prolific's current sample size of native French listeners currently living in France (n= 271) made the desired sample size in the current study logistically impossible. This sample size was selected based on previous literature (e.g. Pardo, 2010) and considering the design of the present task (3 listeners/talker).

Table 5.1. Distribution of native French listener country of birth.

Country of birth	**Number of raters**
France	163
Canada	21
Belgium	15
Switzerland	4
Democratic Republic of the Congo	3
United Kingdom	3
United States	3
Ireland	2
Morocco	2
Australia	1
Cameroon	1
Haiti	1
Mauritius	1
Monaco	1
Somalia	1

Table 5.2. Distribution of native French listener country of current residence.

Country of current residence	Number of listeners
France	136
Canada	24
United Kingdom and Ireland	24
Belgium	17
Portugal	3
Spain	3
Australia	2
Chili	2
South Africa	2
Switzerland	2
United States	2
Germany	1
Greece	1
Finland	1
Norway	1
Poland	1

5.1.2 Materials

The materials used for the native listener judgement task consisted of the items that were collected during the pre-and post-test sessions of the shadowing task, including non-stimulus items. The second repetition of each participant's shadowing of each item was extracted unless it was compromised in some way (produced with abnormal intonation, had additional noises such as coughs, etc.). In these cases, the third repetition was extracted. All words were extracted at zero-crossings in Praat and were normalized for amplitude. In addition to original participant shadowings, the isolated words produced by the model talker for the shadowing task were also used during the judgement task.

5.1.3 Procedures

To determine if film exposure had an effect on global characteristics of French productions, an AXB-style task was implemented that measured perceptual similarity between the model talker and shadowed productions. The goal of this task was to determine whether native French listeners found pre- or post-exposure productions to be more similar to the model talker and therefore, indirectly, more representative of target-like French pronunciation. During each trial, a listener heard three realizations of the same word, one after another. The middle one, item X was the model talker's production of the word, and the flanking ones, items A and B, were the pre- and post-test versions of the same item. Each listener was asked to select whether item A or item B was a better rendition of the item X in the middle, which they indicated by pressing F or J on their keyboard. A and B were counterbalanced for order of presentation and each item was presented four times total (twice in each order). Items within trial were separated by a 200 ms ISI and the following trial began 500 ms after a listener provided a response (following Pardo et al., 2010), otherwise the procedure automatically advanced to the next trial after 5,000 ms elapsed. To keep the task duration manageable, each listener evaluated a single participant. Each participants' productions were presented to 3 different listeners. The experiment was implemented using Gorilla (Anwyl-Irvine et al., 2019) and all instructions/procedures were in French. All listeners were required to complete the study on a computer/laptop and were restricted from participating in the experiment on mobile devices. The experimental task took around six minutes to complete.

Prior to completing the experimental task, all listeners participated in a short headphone screener (Woods et al., 2017), based on dichotic pitch, with the purpose of verifying that all listeners completed the task with headphones. During the headphone screener, a listener heard three pure tones in a row and were asked to identify which tone (A, B, or C) was the quietest. During stimuli presentation, one tone was presented across stereo channels 180° out of phase, making it quite easy to identify the quietest when using headphones, but challenging to distinguish over speakers because of phase cancellation. The screener consisted of six trials and took listeners approximately 2-3 minutes to complete. If a listener selected the incorrect tone on two or more of the trials, they were presented with the screener a second time (including instructions). If a listener failed the headphone screener a second time, they were subsequently rejected from the experiment without completing the experimental task.

After completing the headphone screener, all listeners also completed an identical short practice task in order to get used to the structure of the experiment. During the practice trial, which was identical across all listeners, A and B were produced by talker #23, whose data was not used in any of the analyses due to previous experience with the French language (see section 3.1 for details). The practice task consisted of ten trials and took listeners approximately a minute to complete. Following the practice task, listeners were presented with the instructions one more time and then proceeded to the experimental task.

During the experimental task, data quality was ensured with the inclusion of four attention check items presented randomly throughout the experimental task. These trials were identical in structure to the task, however instead of hearing three pronunciations of the same word, A or B was a different word than the other two. The words used in the attention checks, *canne* (/kan/) and *faune* (/fon/), were selected because they are quite phonetically distinct and would not be easily confused by native French listeners. These items were also not used in any analysis, so additional exposure to the words would not confound the results. Prior to the task, participants were notified of the inclusion of four attention checks and if a listener missed two or more of the attention check questions, their data was excluded from the analysis.

Following completion of the experimental task, listeners took a short background questionnaire (Appendix F), which included the ability to provide the researchers with comments about the study and asked listeners to rate how challenging the experiment was. Using a Likert scale with 1 being “Very easy” and 7 being “Very challenging”, the mean rating for task difficulty across listeners was 4.27 (SD= 1.54).

Prior to beginning data collection, the experiment was tested by two native speakers of French. A total of 33 listeners who completed the study were not included in in the analysis (see Table 5.3) and subsequently replaced with other listeners. The majority of the rejected listeners were those who failed the headphone screener twice, followed by those who missed two or more of the attention checks. Data from three listeners was excluded because they noted in their background questionnaire that they were not native speakers of French despite indicating that they were on Prolific. Two listeners were not included because they skipped five or more experimental trials in a row and one listener was excluded because they did not respond to 15% or more of the experimental trials. All of these exclusions were implemented to ensure the quality of data, which

might otherwise be challenging given the online setting. All excluded listeners were replaced with additional listeners until the desired sample size (n= 222) was achieved.

Table 5.3. Reason for listener data exclusion and exclusion rates.

Reason for exclusion	**Exclusion rate**
Listener did not pass the headphone screener	7.06% (18/255)
Listener missed 2 or more of the attention checks	3.53% (9/255)
Listener was not a native speaker of French	1.12% (3/255)
Participant did not respond to 5 or more experimental trials in a row	0.78% (2/255)
Participant did not respond to 15% or more of the experimental trials	0.04% (1/255)

5.1.4 Analysis

The responses were analyzed with a mixed effects binomial logistic regression using the lme4 package (Bates et al., 2015) in R version 1.2.5033 (R Core Team, 2019). Listener Response (item A or B, corresponding to pre- or post-test participant recordings) was submitted as the binary dependent variable with Participant Group (Audio-Only, Subtitled, and Control) as a fixed effect. Shadower, Item, and Listener variables were included in the model's random effects structure as intercepts and the Control group was submitted as the reference category[10]. Perceptual judgement data was separated into four models. Two models were implemented to analyze French rounded vowel shadowing: one for items containing /y/ and another for items containing /u/. Two additional binomial logistic regression models analyzed French consonant voicing, with one model for items containing voiceless /p/ and another for voiced /b/. If the model in question revealed significant effects/interactions, post hoc pairwise comparisons (Tukey) were used to determine if there were any significant differences between experimental groups.

[10] Similar to the establishment of a random effects structure for the shadowing experiment, a maximal approach with random intercepts for Listener, Talker, Item, as well as random slopes for Group and Session were included in the initial model (Barr et al., 2013). The maximal model that did not run into convergence errors (random intercepts for Shadower, Listener and Item) was ultimately selected.

5.2 Results

A total of 23,088 trials were presented across all native French listeners. Out of all trials, 88 were not responded to (0.38%) by listeners, resulting in a total data set containing 23,000 observations. In preparing data from the native listener perceptual judgement task for statistical analysis, data was first checked for outliers. Responses that were associated with reaction times three standard deviations or more away from each individual listener's mean reaction time were considered outliers and were excluded from any statistical analysis (Jiang, 2013). The elimination of outliers was performed in order to better represent the cognitive processes of the participants and to improve overall data quality. A total of 368 data points (1.60% of the original data set) were omitted, resulting in a total of 22,632 observations in the final data set. The distribution of the remaining data based on the variables in question is presented in Table 5.4.

Table 5.4. Distribution of observations based on variables in question. Note that some tokens represented multiple variables (see Chapter 4.1.2 for details).

Variable in observation		**Total number of observations**
Rounded vowels	/y/	5,227
	/u/	5,240
Word-initial bilabial consonants (VOT)	/p/	5,195
	/b/	5,226
Other tokens (not analyzed)		7,853

5.2.1 Judgements of trials containing rounded vowels

A total of 10,467 judgments (Listener responses) to trials containing either rounded vowel /y/ (5,227 observations) or /u/ (5,240 observations) were submitted to statistical analysis. The first model evaluated whether film exposure had an effect on participants' shadowing of words containing /y/ as assessed by native French listeners (Table 5.5). Results of a mixed effects binomial logistic regression revealed that listeners who heard participants from the Audio-Only group were significantly more likely to select post-test items as more similar to model talker's

speech than listeners who heard Control group participants ($p < 0.01$). Similarly, listeners who judged participants in the Subtitled group were significantly more likely to select post-test items as more similar to model's productions than raters of the Control group participants ($p < 0.01$). These results indicate that native listeners found items from the post-test to be more target-like when participants had received film exposure with original French audio. As demonstrated in Figure 5.1, listeners who heard Audio-Only and Subtitled participants favored post-exposure shadowing as more similar to native speaker realizations, while the order of preference was reversed for Control participants. Post-hoc pairwise comparisons did not find any additional between group differences, indicating that the Audio-Only and Subtitled groups ($p = .99$) were judged approximately the same by native French listeners.

Table 5.5. Table of fixed effects for the mixed effects binomial logistic regression of judgement trials containing /y/.

Predictors	**Odds Ratios**	**Std. Error**	**95% CI**	**Statistic (z-value)**	***p*-value**
(Intercept)	1.17	0.11	0.98 – 1.40	1.70	0.089
Audio-Only	0.73	0.08	0.58 – 0.91	-2.84	0.005
Subtitled	0.72	0.08	0.57 – 0.89	-2.95	0.003

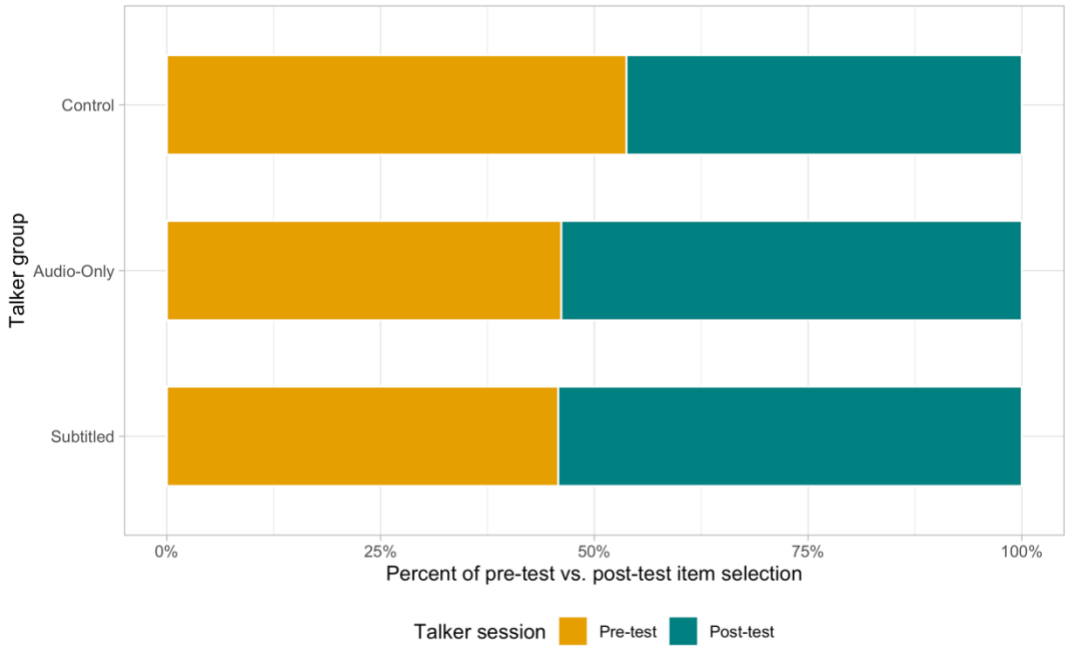

Figure 5.1. Listener selection of talker pre- or post-test /y/ productions as more similar to the model, by Talker group (Control, Audio-Only, Subtitled).

The second model evaluated whether film exposure had an effect on participants' shadowing of words containing /u/ as judged by native French listeners (Table 5.4). Results indicated no significant differences in the log odds of selecting pre- or post-test items as more similar to the model's speech between listeners who attended to different participant groups (Audio-Only vs. Controls: p= .23; Subtitled vs. Controls: p= .61). Therefore, based on native listener judgements, shadowing of /u/ did not change as a function of film exposure. As demonstrated in Figure 5.2, listeners selected pre- and post-test items as more similar to the model's at relatively equal rates across all participant groups.

Table 5.6. Table of fixed effects for the mixed effects binomial logistic regression of judgement trials containing /u/.

Predictors	**Odds Ratios**	**Std. Error**	**95% CI**	**Statistic (z-value)**	***p*-value**
(Intercept)	1.06	0.08	0.91 – 1.23	0.76	0.448
Audio-Only	0.89	0.08	0.74 – 1.08	-1.19	0.233
Subtitled	0.95	0.09	0.79 – 1.15	-0.51	0.607

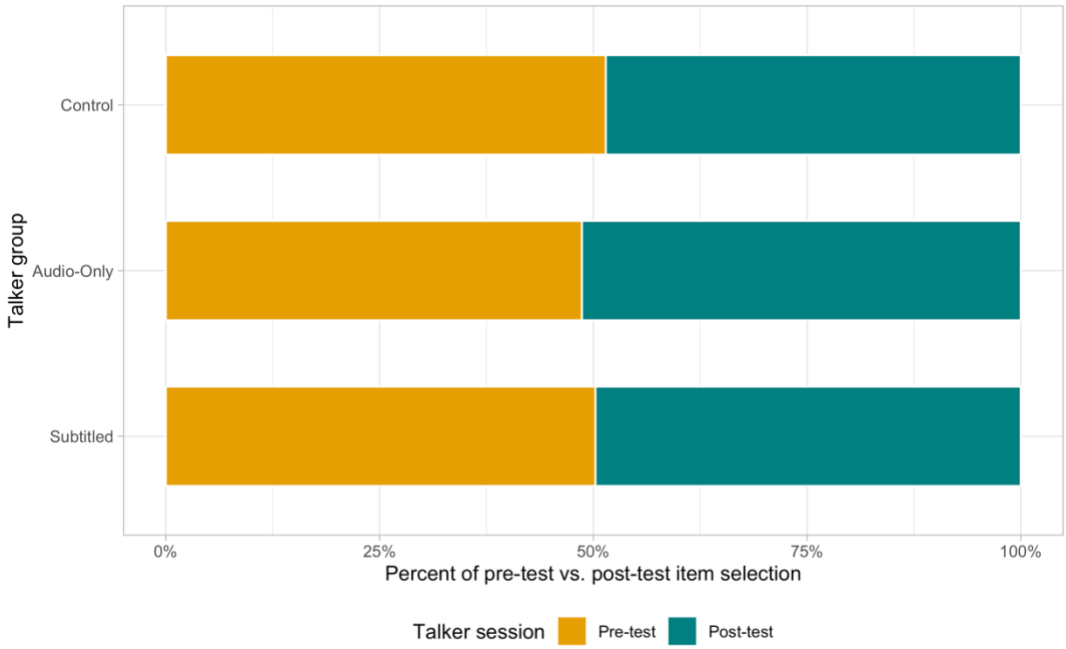

Figure 5.2. Listener selection of talker pre- or post-test /u/ productions as more similar to the model, by talker group (Control, Audio-Only, Subtitled).

5.2.2 Judgements of trials containing word-initial bilabial stops

In order to analyze whether production changes in VOT following film exposure were perceptible to native French listeners, 10,421 observations were submitted to statistical analysis. The first model consisted of trials where listeners heard items containing the voiceless stop /p/ (5,226 observations) (Table 5.7). Results of a mixed effects binomial logistic regression indicated that

listeners who heard talkers from either experimental group (Audio-Only and Subtitled) did not select post-test items at significantly higher rates when compared to listeners who heard Control group talkers (Audio-Only: $p = .11$; Subtitled: $p = .49$). These findings indicate that, according to native listeners, shadowing of /p/ did not significantly change following film exposure (Figure 5.3).

Table 5.7. Table of fixed effects for the mixed effects binomial logistic regression of judgement trials containing /p/.

Predictors	**Odds Ratios**	**Std. Error**	**95% CI**	**Statistic (z-value)**	***p*-value**
(Intercept)	1.03	0.08	0.88 – 1.20	0.32	0.749
Audio-Only	0.86	0.08	0.72 – 1.03	-1.62	0.106
Subtitled	0.94	0.09	0.79 – 1.12	-0.68	0.493

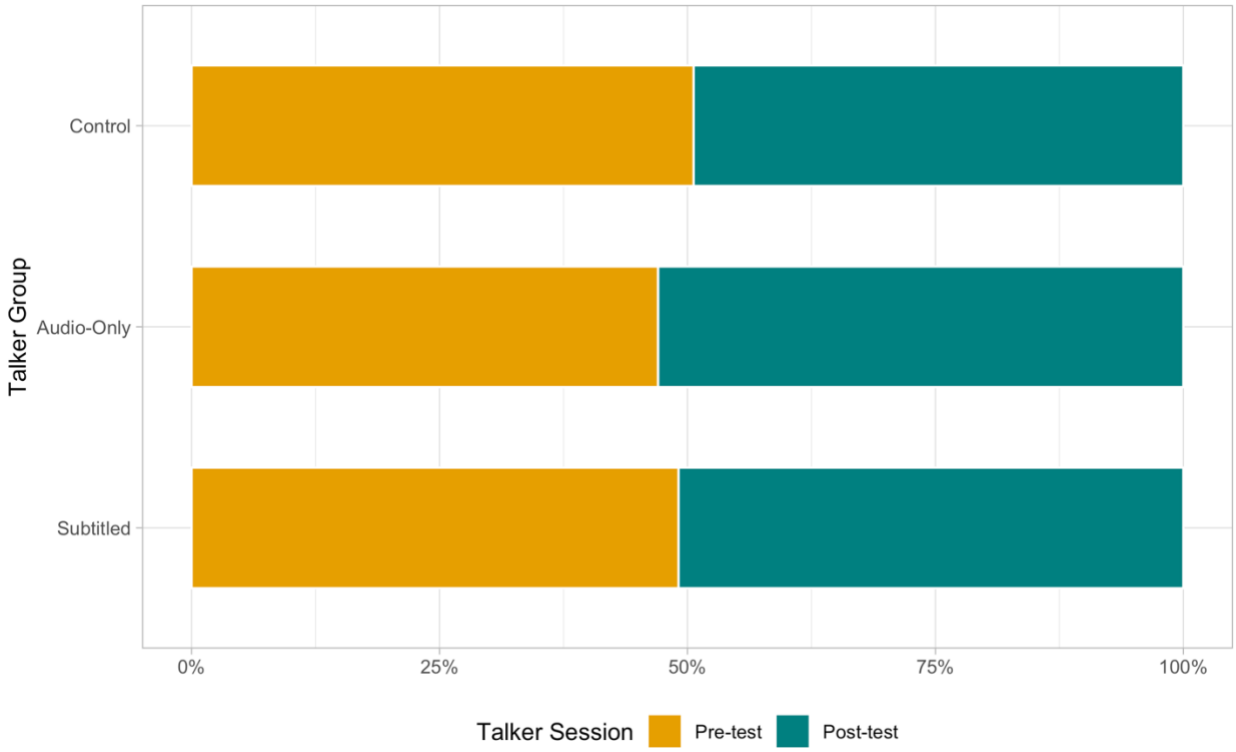

Figure 5.3. Listener selection of talker pre- or post-test /p/ productions as more similar to the model, by talker group (Control, Audio-Only, Subtitled).

Finally, a mixed effects binomial logistic regression was implemented for the 5,195 observations containing /b/ word-initially (Table 5.8). Results that compared selection of pre- or post-test items by listeners who heard the Audio-Only vs. those heard Control group talkers revealed a significant main effect of Group ($p < .05$). This significant effect suggests that native French listeners who heard Audio-Only group talkers were more likely to select post-test productions than those who heard Control group talkers. This finding was also replicated in a comparison of listeners who heard Subtitled group talkers vs. those who heard Control group talkers ($p < .05$). These results suggest that acoustic changes made by the two experimental groups following exposure to foreign film were perceptible to native French listeners (Figure 5.4). Additional pairwise comparisons did not reveal a significant between group difference between the two experimental groups ($p = .99$), indicating that subtitling was not a contributing factor.

Table 5.8. Table of fixed effects for the mixed effects binomial logistic regression of judgement trials containing /b/.

Predictors	**Odds Ratios**	**Std. Error**	**95% CI**	**Statistic (z-value)**	***p*-value**
(Intercept)	1.16	0.10	0.98 – 1.38	1.70	0.089
Audio-Only	0.78	0.08	0.63 – 0.96	-2.29	0.022
Subtitled	0.78	0.08	0.63 – 0.96	-2.36	0.018

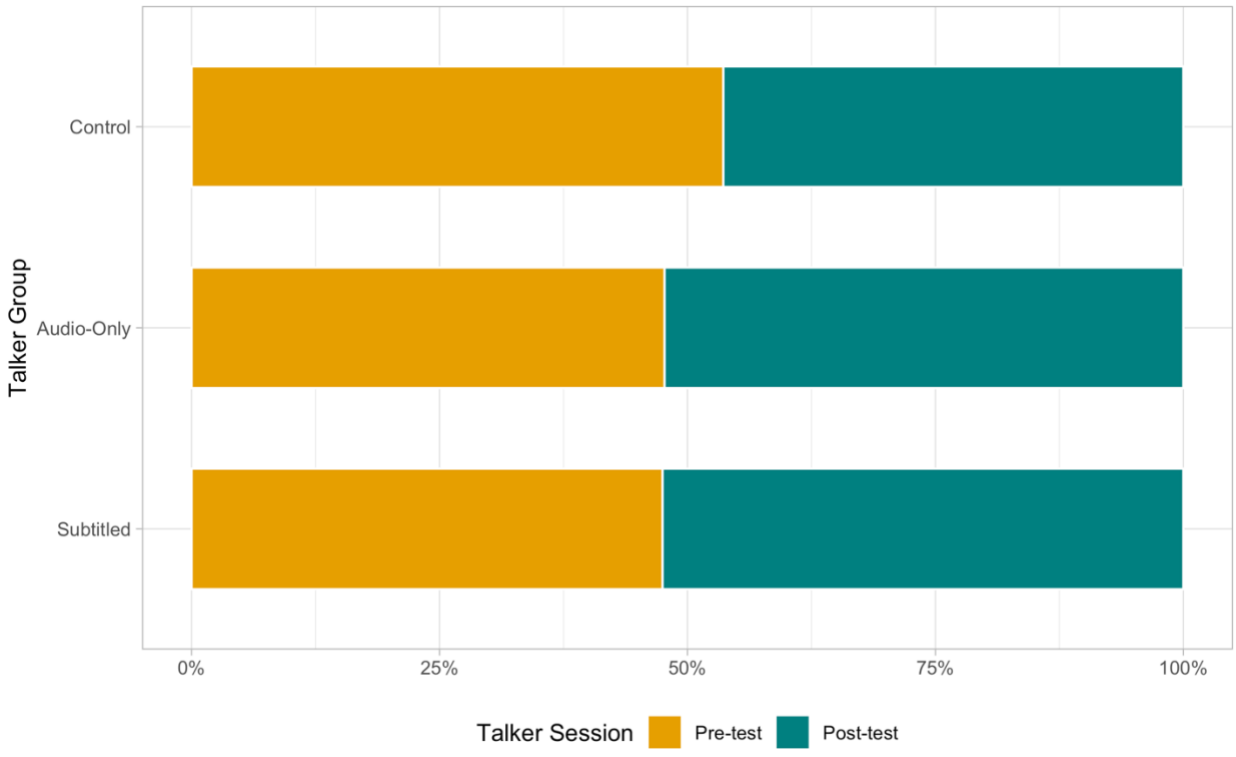

Figure 5.4. Listener selection of talker pre- or post-test productions containing /b/ across each Talker group (Control, Audio-Only, Subtitled).

5.3 Interim discussion

In outlining the research questions and hypotheses for the present dissertation, RQ3 specifically asks whether acoustic modifications to rounded vowels and consonant voicing made by monolingual talkers following exposure to foreign film are perceptible to native French listeners:

(RQ3) Will acoustic modifications to rounded vowels, /y/ and /u/, and consonant voicing made by monolingual American English talkers following foreign film exposure be perceptible to native French listeners?

> Hypothesis 3: If talkers make acoustic changes to their speech following exposure to foreign film, it is hypothesized that these changes will be perceptible to native French listeners. It is also hypothesized that since native listener perceptual judgements provide a more global assessment of phonetic convergence (Pardo et al., 2010), listeners may also pick up on acoustic modifications not analyzed in the acoustic analysis. If this is the case, this will be demonstrated by a higher selection of post-test items in areas where there were not significant acoustic changes revealed in the acoustic analysis.

Regarding rounded vowels, /y/ and /u/, perceptual judgements provided by native French listeners demonstrated that the acoustic adjustments made by experimental group talkers when producing words containing /y/ following film exposure were perceptually salient and selected at higher rates than pre-test productions. In this case, it appears that the significant lowering of /y/ F1 values by both experimental groups following film exposure was likely perceptible by native listeners, providing support for Hypothesis 3. Therefore, though the effect is minor, the role of film is apparent in talker productions of /y/ and it is possible that this effect would increase as film exposure frequency increased.

With regard to perceptual judgements of /u/, while descriptive results demonstrate that listeners who heard Audio-Only group talkers selected post-test items at higher rates than listeners who heard talkers from the other two groups, this result was not statistically confirmed. That being said, despite the Audio-Only group becoming significantly more target-like in their shadowing of F1 following film exposure, these changes were not apparent to native listeners. Therefore, in this case, there does not appear to be support for Hypothesis 3, as acoustic modifications made by talkers following film exposure were not perceptible to native listeners.

One question that arises is why significant F1 value adjustments were enough to clue native listeners into a change following film exposure for /y/, but not for /u/. These perceptual differences are likely due to acoustic modifications that were not captured in the present analysis. That is, the difference in native listener perception of /y/ and /u/, could be that listeners were not paying attention to changes in F1 at all, but instead were focusing on an alternative acoustic attribute(s). As the ultimate goal of the task is to provide a gauge for global improvement, it is impossible to determine what each individual listener was focusing on. If listeners were attuning to a different acoustic adjustment to determine how well the talker converged with the model talker, this might explain the contrasting findings between /y/ and /u/. Given the limits of the current analysis, it is challenging to confirm which of these explanations might be at play in the present experiment. Further investigation and research would need to be completed in order to make definitive determinations regarding the present findings.

Results from an analysis of trials containing voiceless stops, suggest that while there was quantitative difference in the average rate of selection of post-test items by native French listeners who heard experimental group items (especially those who heard Audio-Only group items) vs. those who heard items from the Control group, none of these differences were statistically significant.

Finally, an analysis of trials containing /b/ provided evidence that changes made by both experimental groups following film exposure were perceptible to native French listeners. Based on the acoustic analysis and discussion presented in Chapter 4, it is unlikely that the source was the significant difference in the prevoicing duration following film exposure. If this were the case, it is likely that only the Audio-Only group would demonstrate the effect on listener judgements (they were the only group to significantly change prevoicing durations following film exposure). There is also no evidence from previous research to demonstrate that duration of prevoicing is relevant in categorization of French voicing (Kirby & Ladd, 2015; Netelenbos et al., 2016). It is instead likely that both experimental group talkers were making modifications to something other than VOT when producing /b/ (as suggested in Hypothesis 3) and that those modifications are evident to native French listeners. While VOT is considered primary correlate of voicing, there are many additional acoustic correlates that can contribute to creating a voicing distinction (Lisker, 1986; Llanos et al., 2013), including onset f0 (Abramson & Lisker, 1965; Hombert, 1976; House & Fairbanks, 1953; Kingston & Diehl, 1994; Ohde, 1984), amplitude of aspiration noise (Repp,

1979), oral closure duration (Keating, 1984) and F1 cutback (Liberman et al., 1958; Stevens & Klatt, 1974). Experimental group talkers in the current study could have modified any one of these correlates (or multiple ones) following film exposure, resulting in a higher selection of post-test items by native French listeners. Further exploration into acoustic modifications made by talkers would need to be completed in order to determine the driving force behind the present effect.

In general, the findings of the current experiment line up well with results of the acoustical analysis presented in Chapter 4. These findings therefore support the idea that the significant acoustic adjustments made by talkers following film exposure were perceptible by native listeners, ultimately providing further evidence for the effectiveness of foreign film exposure for acquisition of non-native speech.

CHAPTER 6. EFFECT OF FILM EXPOSURE ON SPEECH PERCEPTION

The focus of the current chapter is to present an analysis of the role of foreign film in non-native speech perception. That is, this chapter discusses how monolingual speakers of English perceive non-native French speech prior to and following exposure to French film. Following the structure of Chapter 4, which presents an analysis of foreign film exposure on French high rounded vowels, /y/ and /u/, and consonant voicing, the current chapter presents an investigation into the same variables with perceptual tasks targeted to address each of them. The first section (Section 6.1) details the cross-language perceptual assimilation task that was designed to examine high rounded vowel perception. Section 6.1.1 presents methods specific to the task, including materials, procedures, data processing and analysis. Section 6.1.2 presents the results and is divided based on vowel, /y/ or /u/. This section concludes with a brief summary of results (section 6.1.3).

Section 6.2 outlines the voicing identification task, including detailing the methods relevant to the task (section 6.2.1). Section 6.2.2 presents the results of the identification task and is divided into two parts: a perceptual identification analysis and a French VOT boundary analysis. A summary of results for the task is presented in Section 6.2.3.

Section 6.3 consists of an interim discussion that breaks down the effect of foreign film exposure on rounded vowel perception (section 6.3.1) and voicing perception (section 6.3.2). This chapter concludes with a section that discusses the perceptual results as a whole (section 6.3.3).

6.1 Perceptual assimilation task

6.1.1 Methods

Participants

The cross-language perceptual assimilation task was administered to the same participants who completed the shadowing task, with the exception of two who were unable to complete the perception task (n=72) due to technical errors. For information regarding participants, see Chapter 3.1.

Materials

All materials used in the perceptual assimilation task were recorded by the same native French model talker who provided the shadowing stimuli, as outlined in Chapter 3. The model talker recorded a list of eight French vowels in the sentence: "La première voyelle du mot *CVC/CV* se prononce *V*" (The first vowel of the word *CVC/CV* is pronounced *V*). The recording session took place in a sound-attenuated booth and each sentence was repeated three times by the model talker into a Shure KSM32 Embossed Single Diaphragm microphone connected to an ART Tube MP Project Series preamplifier. Audio was captured using Audacity version 2.3.2 and digitized at a 44.1 sampling rate.

Vowels were extracted at zero-crossings from the second iteration of the sentence. Extraction from the word containing the target vowel rather than the single vowel was done in order to ensure a more natural pronunciation (Tyler et al., 2014). All stimuli were normalized for intensity and duration using Praat. All vowels were about 200 ms in duration. Prior to administering the task to participants, all stimuli were verified by a different native speaker of French who indicated that they were good representations of the intended categories.

Procedures

In order to determine how exposure to non-native speech via film affected the perception of rounded vowels, a cross-language perceptual assimilation task was conducted (Best, 1995). During each trial, participants were played a single vowel token twice. Following the first presentation, they were asked to indicate which English sound the stimulus resembled the most, using a list of provided English keywords. After selection, the vowel was replayed, and participants rated the similarity of the vowel they heard to the keyword they chose on a continuous analogue scale of 1-5 (1-very unlike the vowel in the English keyword to 5-identical to vowel in the English keyword). The English keywords provided to participants were the following: *heed, hid, aid, Ed, add, odd, awed, hoed, hood, dud, food,* and *heard*. In addition to vowels /y/ and /u/, French vowels (/a/, /u/, /i/, /e/, /o/, /œ/) were also included in the task as fillers.

Prior to the experimental task, participants completed a short practice session that was identical in structure to the experimental session. English vowels were utilized for the practice trial and were recorded by a male, native speaker of English with a language background similar to the

that of the participants. An English practice session was included due to the complex nature of the task, while also giving participants the opportunity to adjust the headphone volume.

Stimuli were presented through Sennheiser HD 380 pro headphones and participants were situated in a quiet room, surrounded by a sound dampening tri-fold stall. The experiment was designed using PsychoPy and participants used a mouse to select their response choice and category goodness rating. Stimuli were randomized and each vowel was presented to participants six times per session (pre- and post-exposure). The task (including the practice session) took participants approximately 7-10 minutes to complete. Following the pre-test session, participants received film exposure and were then administered the same task a second time (post-test). All participants completed the shadowing task first, followed immediately by the perceptual assimilation task.

Data processing and analysis

Data from the assimilation portion of the perceptual assimilation task was analyzed for changes in English response type by group in the pre- and post-tests. A log-linear regression was fit to the data in order to determine whether patterns of perceptual assimilation differed by group or by session. The modal choice was selected as the reference level in the linear regression and the Control group was the Intercept (i.e both experimental groups were compared to the Control group). The English response category was submitted to statistical analysis only if it were selected at least 10% of the time in at least one group (Levy, 2009a). All response categories that did not reach 10% or more were combined into the 'other' category for statistical analysis and data visualization purposes. English response counts across each group and session were submitted as the dependent variable to each model with Group (Control, Audio-Only, and Subtitled), Session, English response, and a Group by Session interaction as predictor variables. A total of two models were run: one for /y/ and one for /u/. The log-linear regression was implemented in R with the following formula: glm(counts ~ session + group + response +group*session, data= DF, family= poisson).

Category goodness ratings were also examined for significant interactions between group and session, and a separate model was implemented for each response vowel. Goodness ratings were analyzed using a linear mixed effects model in R using the lme4 package (Bates et al., 2015). Raw rating score was submitted to the model as the dependent variable, while Group (Audio-Only,

Subtitled, and Control) and Session (Pre- and Post-test) were set as fixed factors and analyzed for interactions. Subject was used as a random effect in the mixed effects model and the Control group was analyzed as the intercept. A t-value > |2.00| was used to determine significance.

6.1.2 Results

Seventy-two participants were administered the perceptual assimilation task and categorized/rated 42 tokens (7 vowels x 6 trials/vowel) in each session (pre- and post-test), which resulted in 6,048 observations in total. Prior to statistical analysis, the data was checked for instances where participants took an unusually long amount of time to provide a response or rating. These outliers were determined by calculating each individual participant's mean reaction time for both responses and ratings. Response/rating reaction times that were three standard deviations away from the individual mean were excluded from statistical analysis (Jiang, 2013). A total of 219 observations (3.62% of the original data set) were omitted. A total of 108 (1.79% of the original data set) of the outliers were from response reaction times and 111 (1.84% of the original data set) were from rating response times. After outlier processing, the final data set contained a total of 5,829 observations. The distribution of items across vowel is presented in Table 6.1. For the sake of the current study, only response and ratings for French vowels /y/ and /u/ (1,658 tokens) will be statistically analyzed.

Table 6.1. Total number of observations per French vowel by session.

French vowel	Pre-test observations	Post-test observations
/ɑ/	408	426
/e/	399	426
/i/	415	429
/o/	384	389
/œ/	440	455
/u/	418	430
/y/	384	426

Table 6.2, 6.3, and 6.5 display the two French vowels, percent English response categorization, and the mean goodness ranking in the pre-test and post-test for the Control group, Audio-Only group, and Subtitled group, respectively. The following sections report the response and rating descriptives and analyses for the target vowels in question, /y/ and /u/.

Table 6.2. Percent response categorization and goodness rating (in parentheses) of each French vowel stimulus (rows) in each session in terms of American English vowel categories/response keywords (columns) for the Control Group. Percentages in bold indicate the most popular response and italicized percentages represent the second-most selected response (adapted from Tyler et al., 2014).

		/i/ heed	**/ɪ/ hid**	**/e/ aid**	**/ɛ/ ed**	**/æ/ add**	**/ɑ/ odd**	**/ɔ/ awed**	**/o/ hoed**	**/ʊ/ hood**	**/ʌ/ dud**	**/u/ food**	**/ɝ/ heard**	**Avg rating**
/y/	Pre-test	*24.36*	7.69	1.28	5.13		2.56		3.85	**29.49**	6.41	12.82	6.41	
		(3.63)	(2)	(2)	(1.50)		(2)		(1.67)	(2.17)	(1.60)	(3.70)	(1.80)	(2.23)
	Post-test	17.07	2.44	1.22	3.66		2.44	4.88	1.22	*21.95*	12.20	**25.61**	7.32	
		(3.50)	(2)	(3)	(2)		(2.50)	(1.75)	(3)	(2.17)	(1.7)	(3.67)	(2)	(2.48)
/u/	Pre-test								1.25	*10.00*		**88.75**		
									(2)	(3.63)		(4.56)		(2.77)
	Post-test									*10.71*	1.19	**88.10**		
										(2.89)	(2)	(4.59)		(3.16)

Table 6.3. Percent response categorization and goodness rating (in parentheses) of each French vowel stimulus (rows) in each session in terms of American English vowel categories/response keywords (columns) for the Audio-Only Group. Percentages in bold indicate the most popular response and italicized percentages represent the second-most selected response (adapted from Tyler et al., 2014).

		/i/ heed	**/ɪ/ hid**	**/e/ aid**	**/ɛ/ ed**	**/æ/ add**	**/ɑ/ odd**	**/ɔ/ awed**	**/o/ hoed**	**/ʊ/ hood**	**/ʌ/ dud**	**/u/ food**	**/ɝ/ heard**	**Avg rating**
/y/	Pre-test	**28.86**	4.03	0.67	5.37	0.67	3.36	4.03	1.34	8.05	12.08	*28.19*	3.36	
		(3.58)	(2.50)	(2)	(2.13)	(1)	(2.20)	(1.83)	(2)	(2.67)	(2.61)	(3.31)	(2.20)	(2.21)
	Post-test	*26.01*	5.20	1.16	4.05	4.05		2.31	0.58	8.09	9.25	**35.84**	3.47	
		(3.60)	(1.89)	(2)	(2.29)	(1)		(1)	(3)	(2.57)	(3.31)	(3.56)	(1.67)	(2.35)
/u/	Pre-test			0.59			4.14	1.18	2.37	*14.79*	3.55	**73.37**		
				(4)			(3.14)	(2.50)	(3)	(3.16)	(2.50)	(4.23)		(3.08)
	Post-test						3.91	1.68	1.12	*8.38*	6.15	**75.42**		
							(1.71)	(1.67)	(2.50)	(3.60)	(3.36)	(4.27)		(2.85)

Table 6.4. Percent response categorization and goodness rating (in parentheses) of each French vowel stimulus (rows) in each session in terms of American English vowel categories/response keywords (columns) for the Subtitled Group. Percentages in bold indicate the most popular response and italicized percentages represent the second-most selected response (adapted from Tyler et al., 2014).

		/i/ heed	**/ɪ/ hid**	**/e/ aid**	**/ɛ/ ed**	**/æ/ add**	**/ɑ/ odd**	**/ɔ/ awed**	**/o/ hoed**	**/ʊ/ hood**	**/ʌ/ dud**	**/u/ food**	**/ɝ/ heard**	**Avg rating**
/y/	Pre-test	*24.84*	4.46		3.18	0.64	1.91		1.27	12.10	7.64	**42.68**	1.27	
		(3.26)	(1.57)		(2)	(1)	(2)		(3.50)	(2.37)	(2.41)	(3.65)	(1.50)	(2.35)
	Post-test	*25.15*	2.92	1.17	0.58	1.17		0.58		11.11	5.26	**49.71**	2.34	
		(3.28)	(2.40)	(1)	(1)	(1.50)		(1)		(2.53)	(1.78)	(3.60)	(2.25)	(2.03)
/u/	Pre-test						1.18		2.37	*13.61*	2.37	**79.88**	0.59	
							(2)		(2.75)	(2.87)	(2.50)	(4.05)	(1)	(2.54)
	Post-test						1.16			*10.40*	5.78	**82.66**		
							(2)			(2.89)	(2.10)	(4.09)		(2.77)

French /y/

As evident in the response categorization and goodness rating tables, English categorization for French front rounded vowels /y/ was highly varied. The model choice for French /y/ across all three groups was English /u/ (30.99% of all responses in the pre-test and 39.43% in the post-test), followed by /i/ (26.30% of all responses in the pre-test and 23.94% in the post-test). Nevertheless, neither response was overwhelmingly frequent, and not every group selected /u/ or /i/ as their top choices. As a result, highly diverging patterns emerged across groups and sessions, as shown in Figure 6.1.

The Control group primarily categorized French /y/ as English /ʊ/ in the pre-test (29.49%) followed closely by 'other' responses (26.92%) and English /i/ (24.36%). Responses changed in the post-test for the Control group, as they indicated /u/ as their most selected response (25.61%) and 'other' as their second-most frequent response (23.17%).

The Audio-Only group assimilated French /y/ most frequently to /i/ in the pre-test (28.486%) with their second most frequent choice being /u/ (28.18%). These responses were switched in the post-test with /u/ becoming the most frequent choice (35.83%) followed by /i/ (26.01%).

The Subtitled group had the highest selection rates for /u/ in both sessions and appeared to maintain categorization ranks between sessions. /u/ was selected most frequently in both sessions (42.68% in the pre-test and 49.71% in the post-test) followed distantly by /i/ as the second-most selected response in both sessions (24.84% in the pre-test and 25.15% in the post-test).

As Figure 6.1 demonstrates, a consistent cross-session change for all groups was an increase of /u/ responses and a decrease of /ʊ/ responses. In the two experimental groups, the increases in responses categorized as /u/ in the post-test did not reduce the amount of /i/ responses in the post-test (/i/ maintained around 25% for both groups in the pre-test and post-test). Instead, they reduced the number of 'other' responses, which were either no longer selected in the post-test (as is the case for /ɑ/) or were not chosen as frequently (for example /ʊ/ and /ʌ/). This pattern does not emerge in the Control group, as the percentage of responses categorized as /i/ is reduced by about 7% in post-test. There is no other prominent evidence that the two experimental groups followed a pattern of cross-sessional change that distinguished them from the Control group.

All English response categories for French vowel /y/ were submitted to a log-linear regression with the most popular choice, /u/, as the reference level (Table 6.5). The model revealed

that when compared to the Control groups in the pre-test, both experimental groups were independent (Control vs Audio-Only: $p < .001$, Control vs Subtitled: $p < .001$). This finding reveals that the three groups did not perform similarly in the pre-test despite being randomly assigned. As demonstrated in Table 6.5, the log-linear model also demonstrated that the rate in which each response other than /u/ was selected was significantly different than /u/.

With regard to the effect of film exposure on perceptual assimilation of /y/, the log-linear model did not reveal a significant relationship between either experimental group and the Control group in the post-test (i.e. a significant Group by Session interaction), suggesting that the data from each group was not independent from the other groups (Control vs Audio-Only: $p = .61$, Control vs Subtitled: $p = .85$). In this case, the lack of significant interactions indicates that participants in each group were consistent in their vowel categorization across sessions.

Table 6.5. Table of poisson regression statistics analyzing selection of English responses for French /y/ (reference category: /u/).

Predictors	**Incidence Rate Ratios**	**Std. Error**	**95% CI**	**Statistic (z-value)**	***p*-value**
Intercept (Pre-test: Control)	27.64	3.39	[21.56, 34.90]	27.07	**<0.001**
Session	1.05	0.17	[0.77, 1.43]	0.32	0.75
Audio-Only	1.91	0.27	[1.46, 2.52]	4.63	**<0.001**
Subtitled	2.01	0.28	[1.54, 2.65]	5.05	**<0.001**
Response /i/	0.71	0.06	[0.59, 0.85]	-3.78	**<0.001**
Response /ʊ/	0.37	0.04	[0.29, 0.46]	-8.82	**<0.001**
Response /ʌ/	0.24	0.03	[0.19, 0.31]	-10.58	**<0.001**
Response ‘other’	0.51	0.05	[0.41, 0.62]	-6.70	**<0.001**
Post-test: Audio-Only	1.10	0.21	[0.76, 1.61]	0.51	0.61
Post-test: Subtitled	1.04	0.20	[0.71, 1.51]	0.18	0.854

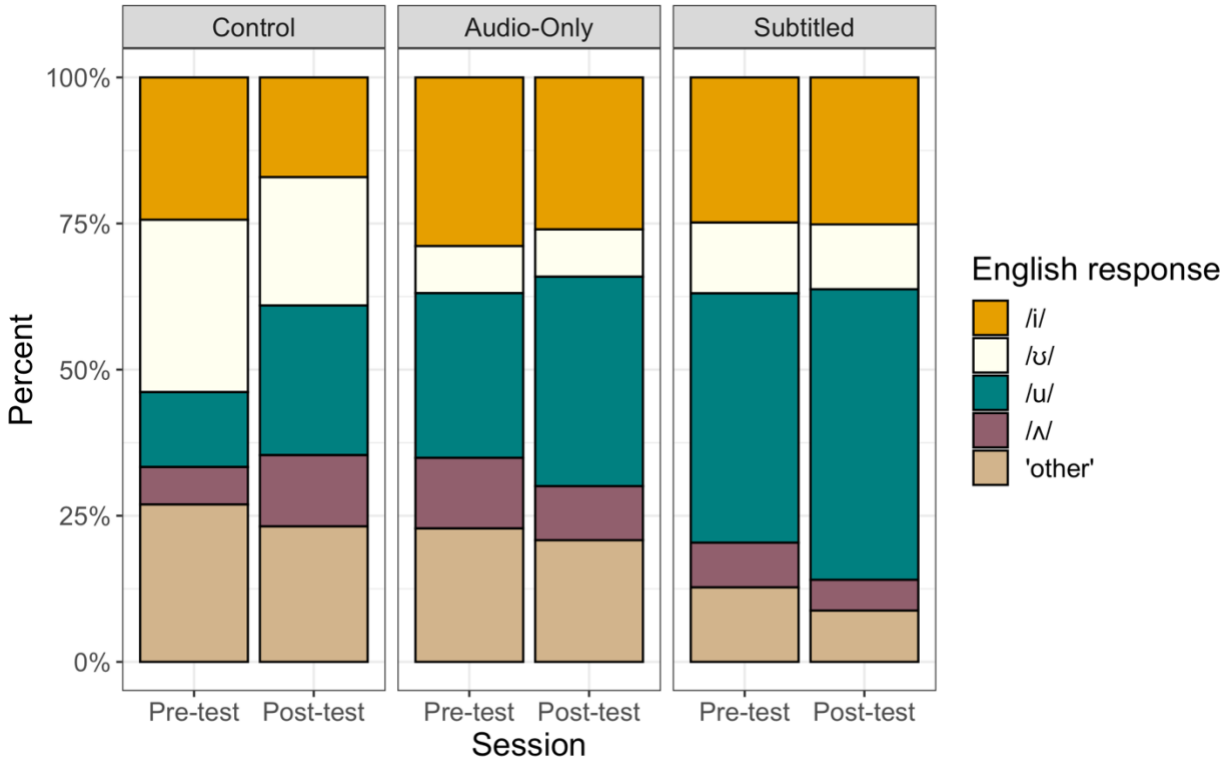

Figure 6.1. Response categorization for French /y/.

French /y/ also received the lowest category goodness ratings out the French stimulus vowels (Pre-test: mean= 2.95, SD= 1.29; Post-test: mean= 3.04, SD= 1.27). This is, perhaps, quite unsurprising, considering the variability in response categorization for French /y/, which suggests that participants were not particularly satisfied with the English response categories they were provided with (Figure 6.2). The most-selected response, /u/, received the highest average category goodness ratings (Pre-test: mean= 3.54, SD= 1.13; Post-test: mean= 3.60, SD= 1.06), followed by the second-most responded category /i/ Pre-test: mean= 3.47, SD= 1.20; Post-test: mean= 3.45, SD= 1.16), which also receives relatively high ratings despite it not being the overall most frequent response. Ratings for /u/ and /i/ are distantly followed by the average ratings for /ʊ/ (Pre-test: mean= 2.35, SD= 0.91; Post-test: mean= 2.41, SD= 1.06) and /ʌ/ (Pre-test: mean= 2.40, SD= 1.24; Post-test: mean= 2.46, SD= 1.17) and then by ratings for 'other' responses (Pre-test: mean= 1.97, SD= 0.99; Post-test: mean= 1.83, SD= 0.94). Of additional interest is that while overall ratings for /u/ were the highest of all response categories, the Audio-Only group rated /i/ higher in both sessions, despite selecting /u/ more frequently in the post-test. Furthermore, when the Control

group selected /ʊ/ as their most frequent English response in the pre-test, it received distinctly lower ratings than /u/ and /i/.

With these results in mind, category goodness ratings for English response choices were submitted to a linear mixed effects model in order to determine if there was an effect of film intervention on response ratings. None of English response categories demonstrated any main effects of Group, Session or any Group by Session interactions (Table 6.6). This result indicates that groups did not behave in markedly different manners, and that there were no pronounced changes across sessions, suggesting that film exposure did not have an effect either.

Table 6.6. Table of fixed effects for the mixed effects model of category goodness ratings for relevant English response categories for French /y/.

Response	**Predictor**	**Estimates**	**Std. Error**	**95% CI**	**Statistic (t-value)**
/i/	Intercept (Pre-test: Control)	3.06	0.44	[2.21, 3.92]	**7.04**
/i/	Audio-Only	0.29	0.54	[-0.76, 1.35]	0.55
/i/	Subtitled	0.09	0.57	[-1.03, 1.21]	0.16
/i/	Post-test	0.18	0.29	[-0.39, 0.76]	0.63
/i/	Post-test: Audio-Only	-0.22	0.33	[-0.87, 0.43]	-0.66
/i/	Post-test: Subtitled	-0.40	0.33	[-1.05, 0.26]	-1.19
/u/	Intercept (Pre-test: Control)	3.51	0.40	[2.73, 4.29]	8.84
/u/	Audio-Only	-0.20	0.46	[-1.11, 0.71]	-0.43
/u/	Subtitled	-0.05	0.45	[-0.93, 0.84]	-0.10
/u/	Post-test	-0.01	0.30	[-0.61, 0.58]	-0.04
/u/	Post-test: Audio-Only	0.23	0.34	[-0.44, 0.90]	0.68
/u/	Post-test: Subtitled	-0.08	0.33	[-0.73, 0.56]	-0.25

Table 6.6 continued

/ʌ/	Intercept (Pre-test: Control)	1.53	0.52	[0.50, 2.55]	**2.92**
/ʌ/	Audio-Only	1.21	0.62	[0.00, 2.41]	1.96
/ʌ/	Subtitled	0.66	0.67	[-0.64, 1.97]	0.99
/ʌ/	Post-test	0.33	0.33	[-0.32, 0.97]	1.00
/ʌ/	Post-test: Audio-Only	-0.61	0.41	[-1.41, 0.19]	-1.50
/ʌ/	Post-test: Subtitled	-0.46	0.47	[-1.38, 0.45]	-0.99
/ʊ/	Intercept (Pre-test: Control)	2.37	0.28	[1.82, 2.92]	**8.46**
/ʊ/	Audio-Only	0.31	0.42	[-0.51, 1.13]	0.75
/ʊ/	Subtitled	0.09	0.38	[-0.65, 0.84]	0.25
/ʊ/	Post-test	-0.02	0.23	[-0.48, 0.44]	-0.08
/ʊ/	Post-test: Audio-Only	0.06	0.43	[-0.78, 0.89]	0.13
/ʊ/	Post-test: Subtitled	0.09	0.37	[-0.65, 0.82]	0.23
‘other’	Intercept (Pre-test: Control)	1.88	0.25	[1.38, 2.38]	**7.43**
‘other’	Audio-Only	0.21	0.32	[-0.41, 0.84]	0.67
‘other’	Subtitled	-0.02	0.35	[-0.71, 0.67]	-0.05
‘other’	Post-test	0.27	0.28	[-0.28, 0.82]	0.96
‘other’	Post-test: Audio-Only	-0.49	0.36	[-1.20, 0.22]	-1.34
‘other’	Post-test: Subtitled	-0.09	0.43	[-0.93, 0.74]	-0.22

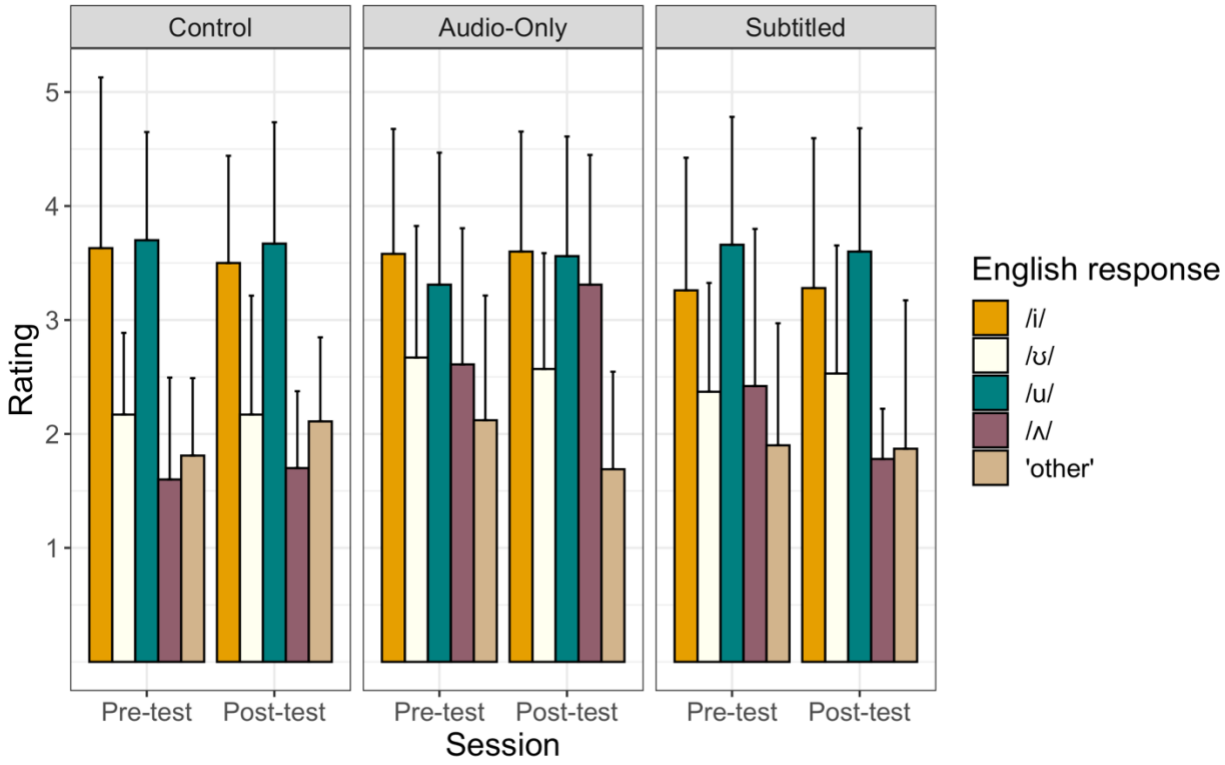

Figure 6.2. Category goodness ratings for French /y/ by English response vowel.

To summarize, French /y/ was assimilated by native English listeners to a wide variety of English vowels, most commonly /u/, /i/, / ʊ /, and /ʌ/, but none of the response categories were judged to be a good representation of the French sound, with the highest rating reaching only approximately 3/5. Despite the great amount of variability across groups and sessions, no consistent pattern of change across sessions distinguished the two experimental groups from the control group, thus not warranting the conclusion that film exposure significantly affected patterns of perceptual assimilation between French and English vowels.

French /u/

High back French /u/ was well established as fitting to English /u/ across all groups and sessions (78.95% of pretest responses and 81.86% of posttest responses), followed by /ʊ/ as the second-most selected response category in all groups (13.39% of pretest responses and 9.77% of posttest responses). The visible pattern demonstrated in Figure 6.3 is the increase in reliance on /u/ and the decrease in /ʊ/ responses between sessions.

As demonstrated in Table 6.7, log-linear regression once again revealed that the groups performed independently of one another in the pre-test, suggesting that despite being randomly assigned there was a significant amount of variation in pre-test English response selection (Control vs Audio-Only: $p < .001$, Control vs Subtitled: $p < .001$). Additionally, when compared to the most popular response, /u/, the other two response categories (/ʊ/ and 'other') were revealed to be significant. This significant effect suggests that these two categories were selected at significantly lower rates compared to /u/ regardless of group or session. That being said, the model did not reveal any significant change in assimilation patterns from pre-test to post-test, including in the experimental groups, thus ruling out the effect of film exposure.

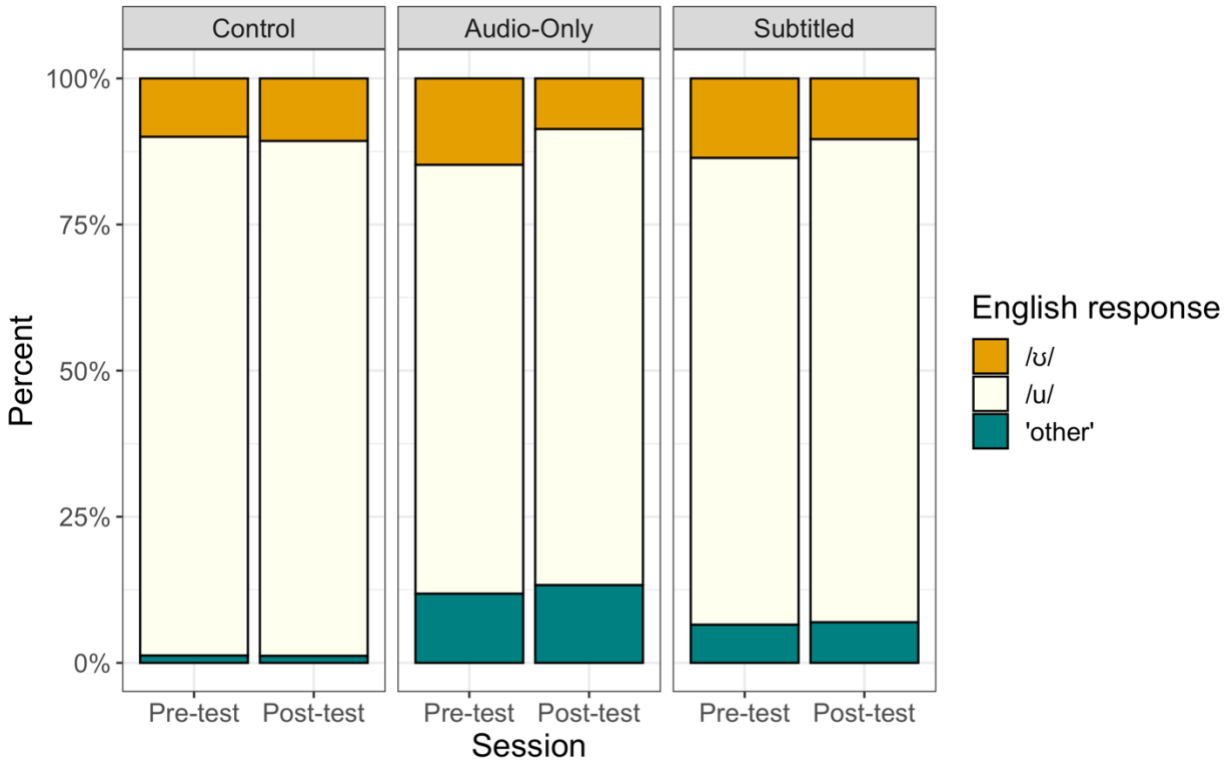

Figure 6.3. Response categorization for French /u/.

Table 6.7. Table of poisson regression statistics analyzing selection of English responses for French /u/ (reference category: /u/).

Predictors	**Incidence Rate Ratios**	**Std. Error**	**95% CI**	**Statistic (z-value)**	***p*-value**
Intercept (Pre-test: Control)	64.34	7.28	[51.13, 79.70]	36.83	**<0.001**
Session	1.05	0.16	[0.77, 1.43]	0.31	0.75
Audio-Only	2.11	0.29	[1.63, 2.77]	5.51	**<0.001**
Subtitled	2.11	0.29	[1.63, 2.77]	5.51	**<0.001**
Response /ʊ/	0.14	0.02	[0.12, 0.18]	-17.96	**<0.001**
Response 'other'	0.10	0.01	[0.08, 0.13]	-18.13	**<0.001**
Post-test: Audio-Only	0.97	0.19	[0.67, 1.42]	-0.13	0.894
Post-test: Subtitled	0.97	0.19	[0.67, 1.42]	-0.13	0.894

Category goodness ratings for French /u/ were fair in the pre-test with an average of 3.96 across all participants. This average rating rose in the post-test to a 4.15 (SD= 1.05). Once again, the highest ratings in both sessions was assigned to the modal response choice, English /u/, which received a pre-test average rating of 4.23 (SD= 0.88) and a post-test average rating of 4.26 (SD= 0.79) across all participants. Second-most choice, English /ʊ/, was also rated fairly highly as a good fit for French /u/, receiving an average rating of 3.11 in the pre-test (SD= 1.11) and 3.14 in the post-test (SD= 1.22). Responses classified as 'other' received the lowest ratings in both sessions with an average of 2.69 in the pre-test (SD= 0.97) and 2.39 in the post-test (SD= 1.08). As Figure 6.4 demonstrates, high category goodness ratings for /u/ were quite stable across

sessions for all groups. Some by-group variability, however, is present for the perceived goodness of fit of English /ʊ/ as a response for French /u/.

A linear mixed effects analysis of category goodness ratings for French /u/ revealed that the Subtitled group rated English /u/ significantly lower than the Control group during the pre-test (t= -2.30), however there was no significant main effect of Session or a Group by Session interaction (Table 6.8). These results indicate that exposure to foreign film did not appear to have an effect on category goodness ratings. Of particular interest is that despite classifying French /u/ as 'other' at higher rates than the other groups, the Audio-Only group did not perceive these options as a better fit for French /u/.

Table 6.8. Table of fixed effects for the mixed effects model of category goodness ratings for relevant English response categories for French /u/.

Response	**Predictors**	**Estimate**	**Std. Error**	**95% CI**	**Statistic (t-value)**
/u/	Intercept (Pre-test: Control)	4.51	0.19	[4.14, 4.89]	**23.60**
/u/	Audio-Only	-0.27	0.24	[-0.73, 0.19]	-1.17
/u/	Subtitled	-0.54	0.23	[-1.00, -0.08]	**-2.30**
/u/	Post-test	0.02	0.09	[-0.15, 0.19]	0.26
/u/	Post-test: Audio-Only	-0.08	0.11	[-0.29, 0.14]	-0.70
/u/	Post-test: Subtitled	0.03	0.11	[-0.18, 0.24]	0.28
/ʊ/	Intercept (Pre-test: Control)	3.39	0.45	[2.50, 4.27]	**7.51**

Table 6.8 continued

/ʊ/	Audio-Only	-0.32	0.52	[-1.35, 0.70]	-0.62
/ʊ/	Subtitled	-0.55	0.53	[-1.59, 0.50]	-1.02
/ʊ/	Post-test	-0.36	0.41	[-1.16, 0.43]	-0.89
/ʊ/	Post-test: Audio-Only	0.85	0.51	[-0.14, 1.85]	1.68
/ʊ/	Post-test: Subtitled	0.25	0.49	[-0.71, 1.20]	0.51
‘other’	Intercept (Pre-test: Control)	2.00	1.06	[-0.08, 4.08]	1.89
‘other’	Audio-Only	0.81	1.10	[-1.34, 2.96]	0.74
‘other’	Subtitled	0.34	1.12	[-1.85, 2.52]	0.30
‘other’	Post-test	0.00	1.50	[-2.94, 2.94]	0.00
‘other’	Post-test: Audio-Only	-0.32	1.52	[-3.30, 2.66]	-0.21
‘other’	Post-test: Subtitled	-0.09	1.54	[-3.11, 2.94]	-0.06

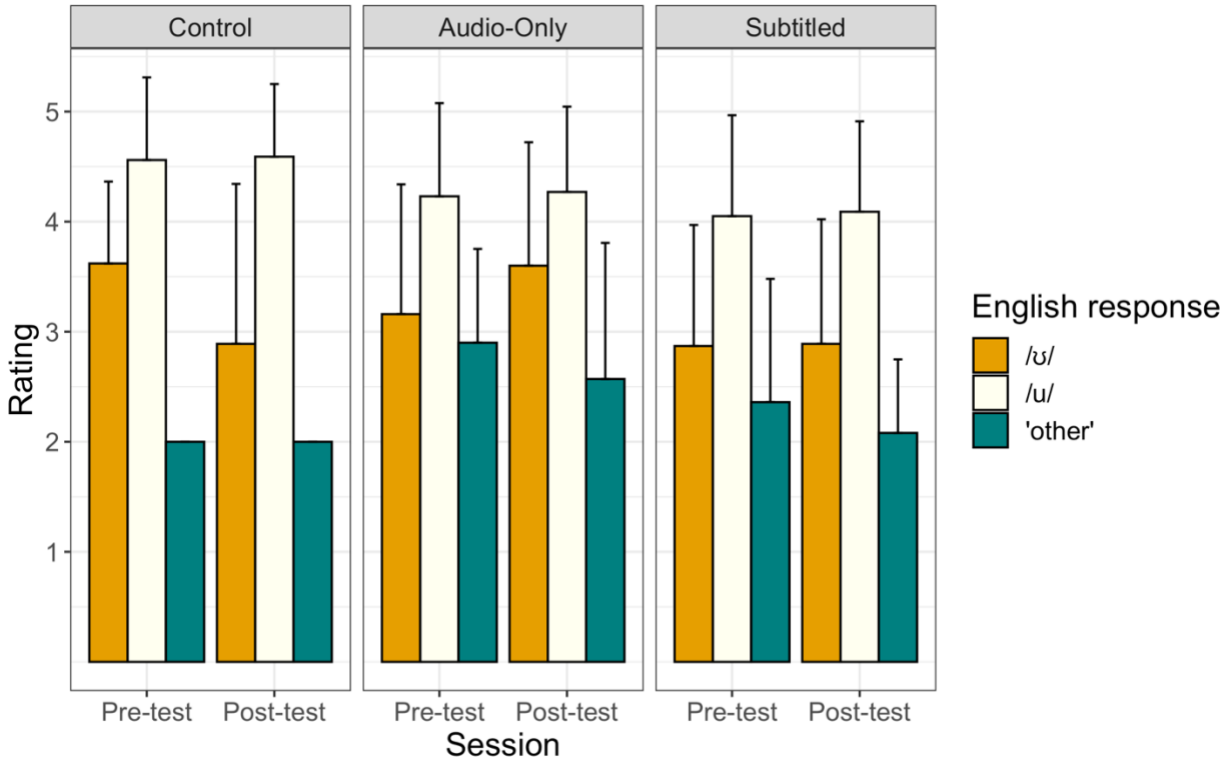

Figure 6.4. Category goodness ratings for French /y/ by English response vowel.

To summarize, unlike /y/, French /u/ was consistently and overwhelmingly assimilated to English /u/ by naïve English-speaking listeners with high degree of confidence (English /u/ was perceived as a good fit for the French /u/). While English /ʊ/ was the second most common choice, it was a considerably less frequent one. Little variability occurred across groups or sessions with respect to the frequency of use of /u/ as the response variable and its rating.

6.1.3 Summary of results

While there was some between-group variation present in the modal choice response for French /y/, in general, the most selected English response across all groups and sessions was English /u/, followed by English /i/. The between-group variation in response selection was confirmed in the multinomial logistic regression with some groups being more likely to select a certain response than other groups. For example, the Control group was more likely to select /ʊ/ instead of modal choice /u/ when compared to the two experimental groups. That said, despite the present variation, results did not provide evidence to suggest that film exposure caused the groups to significantly

alter their response categorization in any way. In line with the lack of response categorization changes, an analysis of category goodness ratings did not reveal any effect of film exposure (i.e. all responses were rated similarly across groups and sessions).

Across groups and sessions, participants considered English /u/ to be a good fit for French /u/, categorizing it to English /u/ at high rates. While between-group variation was present in the response categorization rates (the Audio-Only group perceived French /u/ as an 'other' more frequently than the Control group), this variation was not as wide-spread as in French /y/ categorization. With regard to French /u/, results did not reveal a significant effect of film exposure on either response categorization or category goodness rating, suggesting that foreign film was not effective at altering perception of French /u/.

6.2 Voicing identification task

6.2.1 Methods

Participants

The VOT boundary task was administered to the same participants who completed the shadowing task. For information regarding participants, see Chapter 3.

Materials

In order to create the VOT continuum used in the task, the same model talker who provided recordings for the shadowing task was recorded producing the French words *goût* ("taste") and *cou* ("neck"). The items in question were selected because they are both real words in French (Bond & Fokes, 1991) and did not contain sounds that were unfamiliar to participants (/gu/ and /ku/).

Recordings took place in a sound attenuated booth using an ART Tube MP Project Series preamplifier and a Shure KSM32 Embossed Single-Diaphragm Microphone. Items were digitized at a 44.1 kHz rate. In line with previous literature that applied a similar paradigm (Casillas, 2020; Flege & Eefting, 1986), the voiceless token (/k/), including the following vowel, was resynthesized in Praat. To create the positive end of the continuum, the duration of aspiration was manipulated

until each increment along the continuum was at the desired range. In order to create stimuli with negative VOT, manipulated durations of prevoicing from talker productions of /g/ were pasted prior to the release of stop at zero-crossings. This resynthesis procedure was used to create a VOT continuum that ranged from -110 to 70 ms (following Casillas, 2020). Following resynthesis all tokens were normalized for intensity and vowels were normalized for duration (410.11 ms).

Procedures

During the voicing identification tasks, the experiment was presented to participants using PsychoPy on a desktop computer screen in a quiet room equipped with a sound dampening tri-fold stall. Following instructions, participants were played a single token through a pair of Sennheiser HD 380 pro headphones and were asked to select whether they heard *goût* or *cou* after stimulus presentation. Orthographic labels were presented on the left or right side of the screen and participants indicated their selection by pressing the leftmost or rightmost key, respectively, on a Cedrus response pad (Model RB-740). The placement of orthographic labels on the screen and their corresponding button box keys were counterbalanced across participants. Following each stimulus presentation, a white fixation cross appeared on the screen for a period of 500ms before the next stimulus was presented. Each item was presented to participants once per block and items were presented across eight blocks. All trials were randomized within blocks, and the task took participants approximately four minutes to complete (7-8 minutes across both sessions). Placement of orthographic labels on the screen and their corresponding button box keys were identical across sessions for each participant.

Analysis

In order to assess whether perception of voicing in French changed following film exposure, a series of statistical analyses were completed following procedures outlined in a recent study by Casillas (2020). All statistical analyses were run using the lme4 package (Bates et al., 2015) in R version 1.2.5033 (R Core Team, 2019). First data from the task was submitted to a mixed effects binomial logistic regression. Each model's dependent variable was participant response (/k/ or /g/) and contained fixed effects and interactions for VOT duration, participant Group, and Session. The models' random effects structure included random intercepts for Subject as well as random slopes

for VOT duration and Session, following Casillas (2020). The ultimate goal of these models was to determine if the perceptual identification pattern changed following exposure to French film.

Following this analysis, a series of models was run in order to determine if the perceptual boundaries between voiced and voiceless category changed as a function of Group and Session (i.e. whether film exposure affected the VOT value at which participants crossed over from /g/ to /k/ response). In order to carry out this analysis, the random effects output from the first model was used to calculate each participant's individual boundary crossover point during each session. This calculation was completed using a modified version of the cross_over function from the lingStuff package in R (Casillas, 2018):

$$CO = \frac{\beta_0}{\beta_{VOT}} \times -1$$

In this case, each subject's individual intercepts (β_0) and slopes (β_{VOT}) from the first model were divided by one another and multiplied by -1. These calculations resulted in two total crossover points per participant: Pre-test and post-test. Following calculations, crossover points were submitted to a linear mixed effects model with fixed effects for Group and Session and Group by Session interaction. The random effects structure for the model consisted of random intercept for Subject. T-values were used to determine significance ($t > |2.00|$).

6.2.2 Results

The present experiment that analyzed the effect of foreign film exposure on identification of voicing in French, resulted in a total of 28,000 observations. Following an identical outlier exclusion procedure presented in the perceptual assimilation test analysis, 424 outliers were omitted from the French data set (1.77%). This resulted in a total number of 27,576 French observations.

French perceptual identification

Results from a mixed effects binomial logistic regression that analyzed the effect of foreign film on French perceptual voicing identification (Table 6.9) revealed a significant main effect of VOT

duration that, unsurprisingly, suggests that participant response was dependent on the VOT of the token they heard ($p< 0.001$). An additional VOT by Subtitled group interaction was revealed in the model, which showed that the Subtitled group and the Control group significantly differed in response selection as a function of token VOT duration ($p< 0.05$) regardless of session. However, there was not a present effect of session nor a VOT by Group by Session interaction, which demonstrates that exposure to film did not affect perceptual identification of French voicing.

Table 6.9. Table of fixed effects for the mixed effects binomial logistic regression of French perceptual identification.

Predictors	**Odds Ratios**	**Std. Error**	**95% CI**	**Statistic (z-value)**	***p*-value**
(Intercept)	0.28	0.06	0.18, 0.43	-5.84	**<0.001**
VOT duration	1.08	0.01	1.07, 1.10	16.45	**<0.001**
Audio-Only	0.69	0.19	0.41, 1.19	-1.34	0.181
Subtitled	0.90	0.25	0.53, 1.54	-0.39	0.695
Post-test	0.71	0.14	0.49, 1.05	-1.71	0.087
VOT duration: Audio-Only	1.01	0.01	1.00, 1.02	1.24	0.214
VOT duration: Subtitled	1.02	0.01	1.00, 1.03	2.55	**0.011**
VOT duration: Post-test	1.00	0.01	0.99, 1.02	0.85	0.397
Audio-Only: Post-test	1.22	0.30	0.75, 1.97	0.79	0.427
Subtitled: Post-test	1.11	0.27	0.69, 1.80	0.43	0.665
VOT duration: Audio-Only: Post-test	0.99	0.01	0.98, 1.00	-1.39	0.165
VOT duration: Subtitled: Post-test	0.99	0.01	0.98, 1.01	-1.24	0.216

French VOT boundary analysis

Across groups and sessions, the perceptual boundary between voiced and voiceless stops and French was located at 18.98 ms (SD= 7.20 ms) (average crossover points by group and session

are presented in Table 6.10 and Figure 6.5), a value fairly consistent with an English mode of voicing perception. Of note, is that boundaries appear to be higher in the post-test by about 2 ms in all groups. Results from an analysis of French perceptual boundaries following exposure to foreign film (Table 6.11) did not reveal any effects of Group, which suggests that all groups performed similarly to one another in the pre-test. The model also did not reveal a significant effect of session which would support the notion of a practice effect. Most notably, there was no Group by Session interaction, demonstrating that there was no difference between groups in terms of change across sessions and indicating that participants in the current study did not shift their voicing boundaries following exposure to foreign film.

Table 6.10. Average French crossover points by group and session.

Group	**Session**	**Crossover point (ms)**	
		mean	*SD*
Control group	Pre-test	17.31	7.34
	Post-test	19.52	8.67
Audio-Only group	Pre-test	19.62	6.15
	Post-test	21.84	6.55
Subtitled group	Pre-test	16.39	8.04
	Post-test	18.66	6.58

Table 6.11. Table of fixed effects for the mixed effects model of French perceptual boundaries.

Predictors	**Estimates**	**Std. Error**	**95% CI**	**Statistic (t-value)**
(Intercept)	17.31	1.84	13.72, 20.91	**9.43**
Audio-Only	2.30	2.25	-2.10, 6.71	1.03
Subtitled	-0.93	2.25	-5.33, 3.48	-0.41
Post-test	2.21	1.52	-0.77, 5.18	1.46
Audio-Only: Post-test	0.02	1.86	-3.62, 3.66	0.01
Subtitled: Post-test	0.06	1.86	-3.58, 3.70	0.03

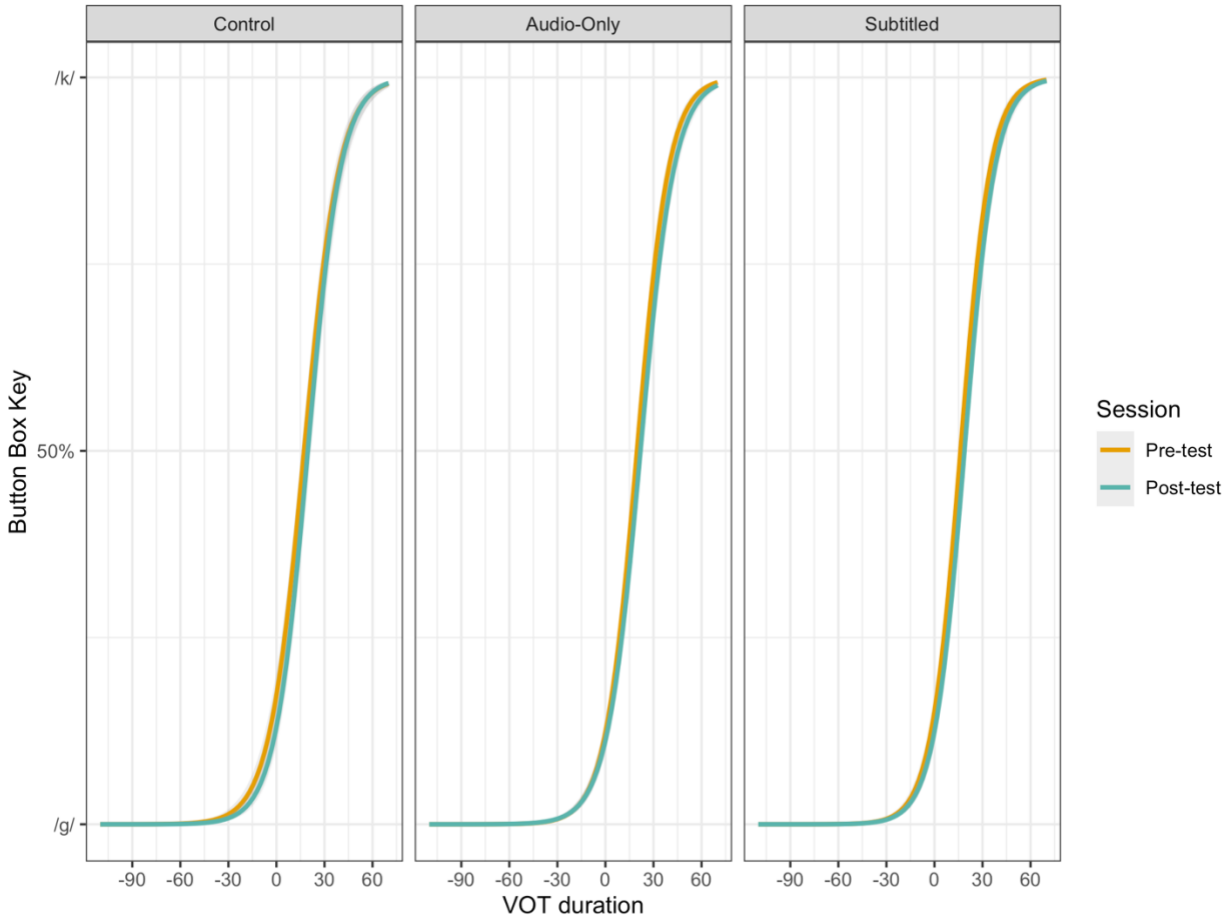

Figure 6.5. Response to French tokens as a function of VOT duration by group and session.

6.2.3 Summary of results

The results of the voicing perceptual identification analysis did not reveal any significant effect of film exposure on selection of voiced and voiceless responses. Additionally, the current results did not provide evidence for an effect of film exposure on French voicing boundaries. That is, the perceptual boundary across all participants (around 18 ms) did not shift in a leftward direction, which would represent a more target-like boundary.

6.3 Interim discussion

The current chapter explores the role of foreign film exposure on the perception of rounded vowels and VOT by implementing a perceptual assimilation task with category goodness ratings and a voicing identification task, respectively. These tasks were specifically selected in order to address

the RQ4 and RQ5, which were presented in Chapter 2. A discussion of these specific research questions with regard to results of the present experiments are addressed in the sections that follow.

6.3.1 Effect of film on rounded vowel perception

In order to examine the effect of foreign film exposure on rounded vowel perception, the following research question and hypothesis was proposed:

(RQ4) Do non-native perceptual assimilation and category goodness ratings of French /y/ and /u/ alter following exposure to foreign film?

> Hypothesis 4: In line with previous literature that applied the same task type (Strange et al., 2009) to a similar population, it is hypothesized that initially all groups will most frequently categorize both French /y/ and /u/ to English /u/. Despite assimilating the two vowels to a single English category, it is predicted that French /u/ will be considered a relatively good fit for English /u/, as demonstrated with category goodness ratings, while /y/ will be considered a poor exemplar of English /u/. Following exposure to foreign film categorization of French /u/ is unlikely to change, however, it is predicted that participants will provide it with lower category goodness ratings after being made implicitly aware of the acoustic differences between the two vowels. With regard to French /y/, while film exposure might affect response categorization, as reflected by higher rates of assimilation to English /i/, ratings are likely to be affected the most. It is hypothesized that exposure to French film will cause participants to become more uncomfortable with their selection, which will be reflected in category goodness ratings.

The results of the perceptual assimilation task with category goodness ratings revealed striking differences in the way that French vowels /y/ and /u/ were perceived by naïve English-speaking listeners (RQ4).

A great amount of variability emerged in assimilation patterns for French /y/, revealing that English listeners were not particularly satisfied with any of the available English vowels provided for categorizing French /y/. As a result, several English vowels (most notably, /u/, /i/, /ʊ/, and /ʌ/) were selected as possible assimilatory categories, but none with very high frequency. Similarly,

goodness of fit ratings were also low across all selected English response types. Another interesting observation was the lack of covariation between the frequency of certain English vowels as categorization responses and their goodness of fit ratings. For example, despite selecting English /ʊ/ more frequently than the experimental groups, the Control group did not rate it significantly higher than the other groups. Most importantly, despite much variability in categorization and goodness of fit ratings across groups and sessions, no consistent pattern of change from pre- to post-test distinguished the two experimental groups from the Control group, suggesting that exposure to non-native speech via film did not significantly affect the way participants perceived /y/ via the prism of the English vowel system. It is possible that the great uncertainly across participants about where French /y/ belongs among English vowels was the reason that there were not trends amidst the variability in responses. A more intensive exposure may be required to produce a consistent effect on these patterns of assimilation.

In contrast to /y/, French /u/ was perceived by participants in a very consistent and confident way, with a low degree of variability across groups and sessions. Specifically, French /u/ was most frequently assimilated to English /u/, which was also judged as a very good fit. Similar to /y/, there was no evidence for the effect of film exposure on the assimilatory patterns of French /u/, ultimately not providing support for Hypothesis 4.

It was predicted in Hypothesis 4 that exposure to French speech via film would decrease participants confidence in their original choices of English vowels as assimilation categories for French sounds. This prediction was not upheld. In fact, no consistent differences in the behaviour of the experimental groups vs. the Control group were revealed. Some significant variability across groups can likely be attributed to combined differences in individual factors affecting non-native speech perception. Variability across sessions most likely reflects a practice effect, since all groups performed the same task twice. In general, the overall complexity and variability of assimilatory patterns may be a reason why the perceptual assimilation task might not be optimal to use when determining the effect of training or other types of exposure on the perception of non-native speech.

Nevertheless, as predicted in Hypothesis 4, results replicated the findings reported in previous literature with respect to assimilation patterns produced by exposure to French /y/ and /u/ in native English speakers, specifically the high variability and low confidence for assimilation of /y/ to English categories like /i/ and /u/ (Gottfried, 1984; Levy, 2009a; Strange et al., 2009), low variability and high confidence in assimilation of French /u/ to English /u/ (Levy, 2009a; Strange

et al., 2009), and partial overlap in assimilatory patterns of the two vowels (both were assimilated to /u/ some of the time) (Levy, 2009a; Strange et al., 2009). Nevertheless, given that /u/ was perceived as a good fit for French /u/, but a poor fit for French /y/ (even when selected as a response), we suspect that the listeners would be able to discriminate between the two vowels, ruling out a 'single category assimilation' scenario (Best, 1995).

6.3.2 Effect of film on VOT perception

In order to address the effect of foreign film exposure on voicing identification and perceptual boundaries, the following research question and hypothesis was proposed:

(RQ5) Do non-native participants alter French perceptual boundaries following exposure to foreign film?

> Hypothesis 5: Casillas (2020) found that over the course of a seven-week immersion program, absolute beginner English learners of Spanish shifted their perceptual boundary in a target-like direction, likely due to a combination of exposure to the target language and explicit instruction. While participants in the current study will not have access to explicit instruction, it is hypothesized that exposure to native input via foreign film will cause them to slightly adjust their perceptual boundary to become more "French-like", as demonstrated by more frequent selection of 'voiceless' that would result in a leftward shift of French perceptual boundaries.

With regard to voicing identification and perceptual boundaries in French (RQ5), it was predicted that experimental group participants would identify tokens more frequently as 'voiceless' in the post-test, resulting in a leftward movement of their perceptual boundaries (Hypothesis 5). These findings would ultimately suggest that exposure to foreign film had caused participants to perceive French VOT in a more target-like manner. However, the current study did not provide evidence to support the aforementioned hypothesis as participants did not alter their selection or boundaries following exposure to foreign film.

Of additional interest is that, in general, the boundary points in which participants crossed over from voiced to voiceless were lower than what has previously been established in English

(crossover is around 30ms; Lisker & Abramson, 1970), but are too high for French, if a 0 ms crossover point is assumed (Caramazza & Yeni-Komshian, 1974; Laguitton, 1997; Serniclaes, 1987). This finding, which is similar to perception of Spanish VOT reported by Casillas (2020), might be due to the phonetic structure of the item used to create the VOT continuum (/ku/). The fact that the resynthesis was based on the voiceless token may have created a general bias to respond with /k/, even at shorter VOT values. In this case, though VOT was resynthesized prior to presentation to participants in the current study, other acoustic correlates were not. That being said, participants in the current study may have been attuning to a correlate other than VOT, ultimately affecting their decisions and leading to the present crossover point.

6.3.3 Exposure to film and non-native speech perception

Results of the present experiment do not provide evidence that film exposure is effective at influencing non-native speech perception for either of the variables examined. With regard to rounded vowels, while task performance might change with an increase in film exposure frequency, it is also possible that the present lack of effects may be due to task type. Previous research has found that how cognitively demanding a perceptual task is can affect results of the task (Best et al., 1996; Flege & Hillenbrand, 1984; Gottfried, 1984; Levy, 2009b; Levy, 2009a; Levy & Strange, 2008; Polka, 1995; Rochet, 1995; Strange et al., 2009). A less cognitively demanding task, like an AX or AXB-style discrimination task, might have been more appropriate in this case considering the naïveté of the participants.

While task type might explain the lack of significant findings for rounded vowel perception, it does not provide an adequate explanation for why there was no change in voicing identification, as identification tasks are relatively straightforward. In this case, as suggested in Chapter 4, it is possible that the lack of perceptual saliency of VOT hindered any changes in perception. In fact, as evidenced by results presented in Chapter 5, it is possible that participants in the present study were not tuning into VOT changes when exposed to foreign film. If this was the case and participants were tuning into other acoustic correlates of voicing, it is not surprising that there were no changes found in the VOT-based voicing identification task. If either of these hypotheses are true, it is possible that perceptual development of VOT is slower and that perception might change with more frequent exposure.

CHAPTER 7. DISCUSSION

Chapter 7 combines results from Chapters 4, 5, and 6 in order to provide a global discussion on the effect of exposure to foreign film on non-native speech acquisition, while also addressing the two specific research questions that remain unanswered (RQ6 and RQ7). Section 7.1 details the relationship between production and perception suggested by the results of the present study, providing potential explanations for some findings reported in previous chapters. Section 7.2 presents a comparison between experimental groups, offering a discussion on which, if either, group experienced a greater effect of film exposure.

7.1 Production/perception relationship

In addition to providing a global discussion on the effect of foreign film exposure on non-native speech acquisition, this section also revisits one remaining research question that addresses the production/perception relationship:

(RQ6) Does exposure to foreign film affect one realm of L2 speech (i.e. production or perception) more than another?

> Hypothesis 6: In line with the general assumption that speech perception precedes production (Levy & Law, 2010), it is hypothesized that talkers will make greater changes in perception following film exposure when compared to production.

Broadly, the goal of the present dissertation was to investigate how exposure to foreign film affects non-native speech production and perception. In general, it was hypothesized that film would have a positive impact on non-native speech production and perception because it serves as a form of HVPT and combines both audio and visual modalities, both of which have been effective training methods in past speech research (Bradlow et al., 1997, 1999; Giannakopoulou et al., 2017; Hardison, 1999, 2003; V. Hazan et al., 2005, 2006; Inceoglu, 2016; Kingston, 2003; Lavan et al., 2019; Lively et al., 1993, 1994; Logan et al., 1991; Ortega-Llebaria et al., 2001; Pruitt et al., 2006; Shen, 2019; Thomson, 2011, 2012; Thomson et al., 2016; Wang et al., 1999, 2003, 2008, 2009; Wiener et al., 2020; Wong, 2012). Additionally, foreign film, which has previously been

demonstrated to affect other realms of L2 acquisition (Dubreil, 2011; Herron et al., 1999, 2000; Kaiser, 2011; Kaiser & Shibahara, 2014; Sherman, 2006; Sturm, 2012; Tognozzi, 2010; Zhang, 2011), provides viewers with access to the authentic native-speaker input necessary for target-like attainment of pronunciation and auditory recognition in the L2 (Flege, 2007, 2018; Flege & Liu, 2001; MacKay et al., 2001).

Results from the production components of the study (acoustic analysis and native listener judgements) suggest that exposure to foreign film had a moderate, but positive effect on talker production. This is evidenced by significant improvements following film exposure in F1 values for /y/ in both experimental groups, which was apparent to native listeners who perceived post-test items with /y/ to be more similar to the model production than pre-test items, in both experimental groups. Additionally, the Audio-Only group produced more target-like F1 values for /u/ following film exposure, though these changes did not affect native listener judgments of similarity to model.

While the acoustic analysis did not reveal any significant post-exposure changes in VOT duration or VOT category frequency, native listeners who heard talkers from both experimental groups perceived post-test items with /b/ to be more similar to model's production than pre-test items, compared to those who heard control group talkers. This finding suggests that although VOT duration/category frequency did not significantly change following film exposure, talkers were making acoustic adjustments elsewhere in order to become more target-like. Together these results provide evidence to support the idea that foreign film exposure has a positive effect on speech production even when exposure sessions are relatively short. It is hypothesized that further exposure could compound already existing benefits.

With regard to speech perception, the current dissertation does not provide evidence that speech perception is affected by exposure to foreign film as demonstrated by participants' performance in a cross-language perceptual assimilation task and the voicing identification task. The combination of production and perception results begs the question of why film exposure was relatively effective at one realm (production), but not the other (perception). While the overall production/perception relationship remains up for debate, some research has suggested that production can precede perception in second language acquisition (Caramazza et al., 1973; Flege, 1991; Flege et al., 1997; Flege & Schmidt, 1995; Flege & Eefting, 1987; Kleber et al., 2012; Mack, 1989; Schertz et al., 2015; Sheldon & Strange, 1982), which could possibly be the case in the current study. This hypothesis is also supported by previous audiovisual training literature (V.

Hazan et al., 2005; Inceoglu, 2016), which found changes in production but not perception following training. If production had the potential to improve earlier than perception for the non-native participants in the present study, film exposure may not have been long enough to affect both realms of speech acquisition. Participants in the current study were only exposed to film during a single, 45-minute session and it is possible that more time is needed to see changes in perception. Further research should be completed to determine if longer and/or more frequent film intervention sessions can alter non-native speech perception.

As mentioned in prior chapters, this interpretation might be exacerbated, or even confounded, by the task types used in the present dissertation. For example, production was measured using an shadowing paradigm, which previous research has suggested does not require phonological categorization (Hao & de Jong, 2016) and is a very simple task. Both perception tasks, however, were quite cognitively demanding for participants and required detailed perceptual assessment and phonological categorization of non-native sounds. If the shadowing task was very simple and/or the perception task cognitively overloaded talkers, it is possible that this influenced their pre- and post-test performance in both modalities (Inceoglu, 2016). This hypothesis is especially likely in the case of /y/-/u/ perception, as task type has previously caused differing results (Best et al., 1996; Flege & Hillenbrand, 1984; Gottfried, 1984; Levy, 2009b; Levy, 2009a; Levy & Strange, 2008; Polka, 1995; Rochet, 1995; Strange et al., 2009). It is therefore possible that a more straightforward, less cognitively demanding task type might have been able to reveal changes in L2 speech perception that the present tasks were unable to reveal. In order to determine if production truly preceded production in this way, an alternative task type(s) should be implemented in future research on film exposure.

Complementing the prior interpretation of why film exposure affected production but not perception, could be the high cognitive demands of foreign film. As demonstrated in Hirata & Kelly (2010), the combination of too many modalities during training might overwhelm participants, hindering their acquisition. This might have specifically affected perception due to the already high cognitive demands of the selected tasks which combined, might have overloaded participants. Once again, the selection of a different perception task in future research might shed more light on what is actually occurring in this case.

An alternate explanation for the results in the present study is that participants were not attuning to the variables that were investigated, leading to discrepancies between production and

perception. For example, an acoustic analysis of VOT duration and category frequency did not reveal any relevant significant changes following exposure to foreign film, but native listeners still selected voiced post-test /b/ items as better realization of model productions than pre-test items at higher rates for experimental group talkers than for control group talkers. This finding suggests that participants were attuning to and altering an acoustic correlate of voicing other than VOT, such as onset f0, amplitude of aspiration noise, F1 cutback, etc. This hypothesis ultimately predicts participant performance on the VOT-based voicing identification task. That is, if they were not attuning specifically to VOT, they would not be expected to alter perception of VOT as a determinant of voicing category, including category boundaries. If this hypothesis is accurate, it also suggests that visual saliency could play a role in which cues/correlates participants are more likely to attune to during film exposure (Inceoglu, 2016). Rounded vowels are more perceptually salient than VOT due to the accompanying visible articulatory gesture, which might have led participants in the present study to tune into the corresponding acoustic characteristics of the vowels. As VOT is not as visually salient, participants might have turned their attention to other acoustic correlates instead, ultimately leading to the difference in outcomes between dependent variables in the present study. Further research on the factor of visual salience and cognitive complexity of the task should be completed in order to determine the roles they play.

7.2 Between group comparisons

The final research question asked in Chapter 2 specifically addresses the role of subtitling on non-native production and perception:

(RQ7) Will access to subtitling during foreign film exposure significantly affect non-native production/perception?

> Hypothesis 7: Although both experimental groups in the current study will receive audiovisual components of the film, it is possible that group viewing the film with English subtitles (Subtitled group) will focus exclusively on the subtitles instead of the film's visuals and thus will not have the benefit of attending to facial gestures. This is further supported by previous research on L2 film, which suggests that reading subtitles requires more cognitive resources in processing and prioritizes visual input over audio (Danan,

1992). As watching a film in a language that talkers are not familiar already has the potential to cognitively overload talkers, it is possible that the additional modality might hinder speech development (similar to the inclusion of a hand modality in Hirata & Kelly, 2010). With the previous research in mind, it is hypothesized that the experimental group that doesn't receive subtitles will out-perform those who do.

With regard to RQ7, while none of the pairwise comparisons examined throughout the dissertation revealed a significant difference between the two experimental groups, quantitative trends in the analyses suggest that the Audio-Only group possibly has a slight advantage over the Subtitled group as predicted in Hypothesis 7. This is supported, for example, by the fact that Audio-Only was the only group to significantly change F1 values of /u/ in the post-test. While the Audio-Only group did not have access to the meaning of the speech they were exposed to in the film, the absence of subtitles and addition of the audio-focused While-Watching activity likely encouraged them to focus on acoustic input and visual gestures. The Subtitled group, however, likely focused primarily on reading, which slightly hindered their access to acoustic and visual information. Previous research on audiovisual pronunciation training (Hirata & Kelly, 2010), has also demonstrated that additional modalities on top of audio and facial input can impede speech learning. It is possible that the addition of subtitles overwhelmed the group (Danan, 1992), and they were not able to improve at higher rates. That said, while the Audio-Only group appears to be somewhat ahead after in a short 45-minute session, it is possible that longer sessions without access to meaning might likely bore participants (the Audio-Only group already scored the lowest on several of the post-participation questionnaire questions regarding enjoyment of and attention to the episode), causing the Subtitled group to begin to outperform the Audio-only group. These conclusions, however, are based on very minor and tentative evidence and the role of subtitling should be examined in greater detail in future research. When the role of subtitling is investigated further, future research might include the implementation of a third group that receives simplified subtitling, which only highlights the main points of the speech. This condition could combine benefits of both experimental groups: greater attention to the audio without the potential of losing focus because of boredom.

CHAPTER 8. CONCLUSION

The present chapter concludes the dissertation by presenting a summary of the results (section 8.1), limitations of the study (section 8.2), and directions for future research (section 8.3).

8.1 Summary of study

The present dissertation explored the effect of exposure to foreign film on the acquisition of non-native speech production and perception. In order to accomplish this, three groups (two experimental, Audio-Only and Subtitled, and one control) of monolingual participants with no prior French experience (n= 74) were asked to complete production and perception tasks in French before and after watching a short film in French. Two variables were selected for analysis in the production and perception domains: high rounded vowels (/y/ and /u/) and consonant voicing (VOT). Production of these variables was elicited with a shadowing task and evaluation included an acoustic analysis, as well as judgements provided by native French listeners (n= 225) through an AXB perceptual judgement task (Pardo et al., 2010). Perception of high rounded vowels and consonant voicing were examined using a cross-language perceptual assimilation task with category goodness ratings (Best, 1995) and a VOT-based voicing identification task, respectively.

Results from the production portion of the study revealed that experimental group participants made significant changes to French /y/ following film exposure (specifically to F1), while changes in /u/ production (F1) were only present in the Audio-Only group. An analysis of VOT duration and VOT category frequency did not demonstrate any relevant significant changes following exposure to foreign film. Supplementing the acoustic analysis, findings from the native listener judgement task demonstrated that the changes experimental groups made to /y/ following exposure to film were perceptible and made participants sound more like a native French model. Results additionally indicated that participants made positive, perceptible changes to their production of voiced consonants, despite the fact that acoustic analysis did not reveal any post-exposure changes in VOT. This finding suggests that participants were likely attuning to and altering an alternative acoustic correlate of voicing, which caused native listeners to perceive participants' productions to be more target-like.

In comparison, current perception results do not provide evidence that perception of French vowels and consonants was affected by exposure to foreign film.

While significant results in the present dissertation are limited and are restricted to the production domain, they should not be overlooked, as participants in the present study had no prior experience with French and only took part in a single exposure session. Additional research, outlined further in section 8.3, on the effect of film exposure should be completed in order to determine if longer and/or more frequent sessions compound benefits found in the current study.

8.2 Limitations

Although this study demonstrates the effect of foreign film exposure in improving some aspects of non-native speech, it is not without limitation. Notably, there is one main theoretical limitation. While this dissertation hopes to be applicable to L2 speech acquisition, it cannot be directly applied as participants in the present study were not current L2 learners. Further research should be conducted with current L2 learners of the target language to confirm that findings are generalizable (see section 8.3 for details).

There are additionally several practical limitations. Firstly, although the native listener perceptual judgement task (Chapter 5) was divided into four separate analyses, each of which examined a different variable, due to the global nature of the task it is impossible to determine if listeners were actually providing judgements that were dependent on that specific variable. For example, when listening to the word *vu* (/vy/), listeners might have been picking up on changes in vowel quality, but word-initial fricative production could have also informed their judgement. This limitation was further exacerbated as some of the stimuli in the present study occurred in multiple lists (rounded vowels and consonant voicing). For example, the word *bulle* occurred on both lists as it contained /b/ and /y/. Experimental group production of both of these variables was significantly affected by film exposure, but it is challenging to confirm which aspects of words that occurred on both lists were more influential in determining the outcomes of the perceptual judgement task. Therefore, while patterns in the present dissertation suggest that certain variables were informing listener judgements, these results should be approached with caution. Future research should consider using more distinct lists or isolating variables in order better pinpoint what factors are at play.

An additional limitation of the present study, also regarding the AXB task, is that while the population of listeners were all native speakers of French, they were born in and/or currently resided in a variety of places. While the lack of homogeneity among listeners could have led to judgements influenced by dialect-specific features, homogeneity was impossible due to the current population of native French speakers active on Prolific (n= 271). That said, there is currently no research to suggest that dialect-specific features influence native listener perception of phonetic convergence. In fact, the opposite is true. For example, in a study that examined the effect of dialect experience of native speakers of American English on the perception of phonetic convergence by other native speakers of American English, experience with a certain dialect did not lead to significant differences between groups (Ross et al., forthcoming). Though this study is different in a number of ways when compared to the present study (L1 speakers judging other L1 speakers, dialects examined were all those of American English, etc.), it does suggest that dialectal homogeneity may not be entirely necessary when collecting native listener perceptual judgements. It is also noted that even if the study was able to recruit native French listeners born and currently living in France, for example, there would not be a way to control for inter-speaker diversity using an online data-collection platform like Prolific. Finally, as French listener proficiency in English was not controlled, it could have potentially influenced provided judgements. For example, if a listener had higher English proficiency and/or more frequent English input they might have judged the talkers differently from those who had lower English proficiency and/or less frequent English input.

Another limitation of the present study is the lack of a delayed post-test. As a result, it is difficult to tell which, if any, of the improvements could be maintained following exposure to film, and for how long. Although a delayed post-test was originally planned, the logistics of scheduling an additional meeting with 74 participants made it impossible to execute. That said, future research, especially the examination of foreign film exposure with participants who regularly use or are actively learning the language should consider the implementation of a delayed post-test in order to better understand the trajectory of learning.

8.3 Future directions

Broadly, this research could be expanded in two potentially interesting directions, each of which would provide a more developed understanding of the effect of film exposure on non-native speech. Firstly, and perhaps most importantly, longer/more frequent exposure sessions should be implemented in order to determine if this leads to a greater effect in the production and a significant effect in perceptual domains. While previous literature has shown that participants can capitalize on the benefits of a training program within minutes of exposure (Pisoni et al., 1982), foreign film is not designed specifically for the purpose of L2 speech development and its effect could take longer to manifest itself. It is predicted that participants exposed to film for a longer amount of time, whether over a single or multiple sessions, will experience larger gains in the production domain, which could possibly apply to less visually salient variables, like VOT. Additionally, it is hypothesized that longer exposure will cause participants to adjust the perception domain, which was not present in the current dissertation. This is supported by the marginal changes found in perceptual assimilation of /y/ presented in Chapter 6. It is possible that further exposure would cause these changes to become more pronounced.

The second major future direction of this research program would be to apply the current methodology to a group of participants who already have experience with the target language in order to determine if similar or greater effects are found. As previous research supports the positive influence of authentic native speaker input in the L2 on target-like acquisition of L2 speech production and perception (Flege, 2007, 2009, 2018; Flege et al., 1997; Flege et al., 1995; MacKay et al., 2001; among others), it is worth investigating to what extent foreign film can provide a similar benefit to L2 learners. Considering findings in the present study, though current L2 learners would start in a different place due to higher proficiency level, it is hypothesized that they would benefit from film exposure in a similar way to those in the present study (especially if offered longer exposure). This hypothesis is also supported by previous literature that applied equivalent training procedures to both monolingual speakers and L2 learners (Lively et al., 1994; Logan et al., 1991) and demonstrated a similar level of success. While film is not a specially designed training procedure, it does provide viewers with naturalistic exposure to L2 speech, which could lead to similar benefits across different populations.

When considering the expansion of this research program into the aforementioned directions, there are also several aspects of the methodology that could be modified in order to

achieve a more detailed understanding of the role of film in the development of non-native speech. These include the selection of a different task type (perception and/or production), the examination of different variables and language pairings, expansion of between group analysis (for example including an additional experimental group that receives simplified subtitling), and the implementation of a delayed post-test in order to determine if improvements following foreign film exposure are retained over a period of time.

While the results of the present study are by no means definitive, they are strongly suggestive of the fact the greater exposure to foreign film may benefit L2 learners, specifically in the domain of non-native/L2 speech production. That being said, exploration into the aforementioned research directions and the application of any of the suggested modifications would provide a more nuanced understanding of the role of foreign film exposure in non-native/L2 speech. Although further research is needed in order to make definitive conclusions, the incorporation of foreign film into the L2 classroom is likely to have a positive impact on L2 speech development and, more generally, appears to be a useful and effective pedagogical tool.

APPENDIX A. PARTICIPANT LANGUAGE BACKGROUND QUESTIONNAIRE

Instructions: We would first like to ask you some basic questions concerning your background. This is not a test, so there are no right or wrong answers. In order to keep your identity confidential, your responses will be anonymized. All records will refer to your participant identification number only. No record will be linking your identity to specific data. You may choose to skip any question that you do not feel comfortable answering.

1. How old are you?
   Question type: Text entry

2. What is your gender?
   Question type: Multiple choice
   a. Male
   b. Female
   c. Prefer not to say
   d. Other_____

3. What is your sexual orientation?
   Question type: Slider

Exclusively homosexual						Exclusively heterosexual
1	2	3	4	5	6	7

4. What is your primary political orientation?
   Question type: Multiple choice
   a. Liberal
   b. Conservative
   c. Independent
   d. Prefer not to say

e. Other_____

5. Have you ever had a vision problem, hearing impairment, language disability, or learning disability?

   Question type: Multiple choice

   a. Yes
   b. No

6. If yes, please explain (including any corrections)

   Question type: Text entry

7. Where were you born (town, state/region, country)?

   Example: West Lafayette, IN, USA

   Question type: Text entry

8. Where have you lived (give each location and number of years you lived there)?

   Example: West Lafayette, IN- 4 years

   Question type: Text entry

9. Was English the first language you spoke?

   Question type: Multiple choice

   a. Yes
   b. No

10. If no, what age were you when you started speaking English?

    Question type: Text entry

    Note: This question was only displayed if 'No' was selected on question 9

11. If no, what was the first language you spoke

    Question type: Text entry

    Note: This question was only displayed if 'No' was selected on question 9

12. Think of the adults who raised you. Was English their first language?

    Question type: Multiple choice

    a. Yes
    b. No

13. If no, please indicate the language each adult spoke and their relationship to you.

    Example: Father- Japanese, Mother- English

    Question type: Text entry

    Note: This question was only displayed if 'No' was selected on question 12

14. Have you studied any language other than English?

    Question type: Multiple choice

    a. Yes
    b. No

15. If yes, indicate each language, your age when you started studying it, and for how long you studied it.

    Example: Spanish, 7 years old, 10 years studied

    Question type: Text entry

    Note: This question was only displayed if 'Yes' was selected on question 14

16. If yes, rate your proficiency in the language(s) you mentioned, using the following scale:

    1- Very poor
    2- Poor
    3- Fair
    4- Functional
    5- Good
    6- Very good
    7- Native-like

    Question type: Text entry

    Note: This question was only displayed if 'Yes' was selected on question 14

17. Have you ever lived in or visited any place where any language other than English is used for daily communication?

    Question type: Multiple choice

    a. Yes
    b. No

18. If yes, indicate each place, the language(s) spoken there, the age you were when you were visiting/living there, and the amount of time spent there (treat separate trips as distinct entries).

    Example: Switzerland, French & German, 18 years old, 2 weeks

    Question type: Text entry

    Note: This question was only displayed if 'Yes was selected on question 17

19. How much time do you estimate you are around people who speak English as a second language (non-natively)?

    Question type: Matrix table

    a. None at all
    b. Very little
    c. A small amount
    d. A moderate amount
    e. A lot
    f. A great deal
    g. All of the time

20. If you spend any time with people who speak English as a second language (non-natively), please indicate to the best of your ability where those people are from.

    Example: China, France, etc.

    Question type: Text entry

21. How much time do you estimate you watch foreign films (films where the main language is not English)?

    Question type: Matrix table

    a. None at all
    b. Very little
    c. A small amount
    d. A moderate amount
    e. A lot
    f. A great deal
    g. All of the time

22. If there is anything interesting or important about your language background or language use that you would like to share, please comment below.

    Question type: Text entry

APPENDIX B. WHILE WATCHING ACTIVITY FOR EXPERIMENTAL GROUPS

G1G2 Activity

Instructions: You will now watch a short documentary in French. Although you have expressed that you don't know French, we are interested to see how many words you recognize while you watch this episode. If you hear a word that you recognize, record it on this sheet of paper. These words can be French words that you've heard before or French words that sound similar to English words. You do not have to fill up the paper, but please write as many words as you can.

APPENDIX C. WHILE WATCHING ACTIVITY FOR CONTROL GROUP

G3 Activity

Instructions: You will now watch a short documentary in French. We are interested to see how many food-related words (i.e. words like ‘celery’, ‘skillet’, ‘beef’, etc.) you recognize while you watch this episode. If you hear a food-related word that you recognize, record it on this sheet of paper. You do not have to fill up the paper, but please write as many words as you can.

APPENDIX D. POST-PARTICIPATION QUESTIONNAIRE

1. Participant number (to be filled out by the researcher)
   Question type: Text entry

2. Prior to watching Chef's Table: France today, had you ever seen an episode of Chef's Table?
   Question type: Multiple choice
   a. Yes
   b. No

3. If yes, were any of those episodes specifically Chef's Table: France?
   Question type: Multiple choice
   Note: This question was only displayed if 'Yes' was selected on question 2
   a. Yes
   b. No

4. Using the scale below, please indicate how much you like watching food-related shows.
   Question type: Matrix table
   a. Dislike a great deal
   b. Dislike a moderate amount
   c. Dislike a little
   d. Neither like nor dislike
   e. Like a little
   f. Like a moderate amount
   g. Like a great deal

5. Using the scale below, please indicate how much you liked watching the episode of Chef's Table: France **today**.
   Question type: Matrix table

a. Dislike a great deal
b. Dislike a moderate amount
c. Dislike a little
d. Neither like nor dislike
e. Like a little
f. Like a moderate amount
g. Like a great deal

6. Using the scale below, please indicate how much you think you paid attention to the episode of Chef's table: France **today**.
   Question type: Matrix table
   a. None at all
   b. Not much at all
   c. A little bit
   d. A moderate amount
   e. Some
   f. A lot
   g. A great deal

7. Using the scale below, please indicate how much you think you paid attention to the episode of Chef's Table: France's **audio** (i.e. people talking, music, etc.).
   Question type: Matrix table
   a. None at all
   b. Not much at all
   c. A little bit
   d. A moderate amount
   e. Some
   f. A lot
   g. A great deal

8. While you were watching the episode, were subtitles included?

Question type: Multiple choice

a. Yes
b. No

9. Using the scale below, please indicate how much you think you paid attention to the subtitles during the episode.
   Question type: Matrix table
   Note: This question was only displayed if 'Yes' was selected on question 8
   a. None at all
   b. Not much at all
   c. A little bit
   d. A moderate amount
   e. Some
   f. A lot
   g. A great deal

10. Using the sliding scale below, please indicate what percentage you could understand the language the show was in
    Question type: Slider ranging from 0-100.

11. Before and watching Chef's Table: France, you completed a task where you repeated French words after a speaker. Using the scale below, please indicate how effective you believe watching Chef's Table: France was in improving your pronunciation of the French words.
    Question type: Matrix table
    a. None at all effective
    b. Moderately ineffective
    c. Slightly ineffective
    d. Neither effective nor ineffective
    e. Slightly effective
    f. Moderately effective

g. Extremely effective

12. If there are any other comments you'd like to share about your participation in this study, please do so below.
    Question type: Text entry

APPENDIX E. PRE-SHADOWING SHORT TEXT

From the speech accent archive (Weinberger & Kunath, 2011): *Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.*

APPENDIX F. LISTENER LANGUAGE BACKGROUND QUESTIONNAIRE

1. En utilisant l'échelle ci-dessous, veuillez décider de la difficulté de l'activité que vous venez de compléter.
   English translation: Using the scale below, please rate how difficult you found the experimental task you just completed?
   Question type: Slider ranging from Très facile (Very easy) to Très difficile (Very difficult)

2. Quel âge avez-vous ?
   English translation: How old are you?
   Question type: Text entry

3. Quel est votre sexe ou genre actuel ?
   English translation: What is your gender?
   Question type: Text entry

4. Dans quel pays êtes-vous né(e) ?
   English translation: What is your country of birth?
   Question type: Text entry

5. Dans quel pays vivez-vous actuellement ?
   English translation: What country do you currently reside in?
   Question type: Text entry

6. Le français est-il la première langue que vous avez apprise étant enfant ?
   English translation: Was French the first language you learned as a child?
   Question type: Multiple choice
   a. Oui (Yes)
   b. Non (No)

7. Si non, quelle est votre langue natale ?
   English translation: If no, what was the first language you spoke?
   Question type: Text entry

8. À quel âge avez-vous commencé à apprendre l'anglais ?
   English translation: At what age did you begin learning English?
   Question type: Text entry

9. En utilisant l'échelle ci-dessous, veuillez décider de votre niveau d'anglais.
   English translation: Using the scale below, please rate your overall proficiency in English?
   Question type: Slider ranging from Très bas (Very poor) to Presque natif (Native-like)

10. Parlez-vous d'autres langues, en plus de l'anglais et du français ?
    English translation: Do you speak any other languages besides French and English?
    Question type: Multiple choice
    a. Oui (Yes)
    b. Non (No)

11. Si oui, quelle(s) autre(s) langue(s) parlez-vous et à quel âge avez-vous commencé à apprendre ces langues ? (exemple : Allemand, à 3 ans)
    English translation: If yes, what additional languages do you speak and at what age did you begin learning them? (Example: German, 3 years old)
    Question type: Text entry

12. Êtes-vous droitier ou gaucher ?
    English translation: Which hand is your dominant hand?
    Question type: Multiple choice
    a. Droitier (Right)
    b. Gaucher (Left)

13. Avez-vous reçu une instruction musicale (chant et/ou musique) ?

English translation: Have you ever had any formal musical or voice training?

Question type: Multiple choice

a. Oui (Yes)
b. Non (No)

14. Avez-vous déjà eu des problèmes d'audition, prononciation, vision (en dehors d'une simple correction visuelle), ou d'apprentissage ?

    English translation: Have you ever had a vision problem hearing impairment, language disability, or learning disability (not including corrections for vision)?

    Question type: Multiple choice

    c. Oui (Yes)
    d. Non (No)

15. Si oui, veuillez expliquer (incluez les corrections ou traitements effectués).

    English translation: If yes, please explain (including any corrections).

    Question type: Text entry

16. Veuillez entrer votre identifiant Prolific.

    English translation: Please provide us with your Prolific ID.

    Question type: Text entry

REFERENCES

Abramson, A. S., & Lisker, L. (1965). Voice onset time in stop consonants: Acoustic analysis and synthesis. *Proceedings of the 5th International Congress of Acoustics, 51,* A51.

Akahane-Yamada, R., Strange, W., Downs-Pruitt, J. C., & Masuda, Y. (1998). Modification of *L* 2 vowel production by perception training as evaluated by acoustic analysis and native speakers. *The Journal of the Acoustical Society of America, 103*(5), 3089–3089. https://doi.org/10.1121/1.422989

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N. Z., & Evershed, J. K. (2019). Gorilla in our midst: An online behavioural experiment builder. *Behavior Research Methods*. https://doi.org/10.3758/s13428-019-01237-x

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/doi:10.18637/jss.v067.i01

Battye, A., Hintze, M.-A., & Rowlett, P. (2003). *The French language today: A linguistic introduction*. Routledge.

Best, C. T. (1995). A direct realist view of cross-language speech perception. *Speech Perception and Linguistic Experience*, 171–206.

Best, C. T., Faber, A., & Levitt, A. (1996). Assimilation of non-native vowel contrasts to the American English vowel system. *The Journal of the Acoustical Society of America, 99*(4), 2602–2603. https://doi.org/10.1121/1.415316

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language Learning & Language Teaching* (Vol. 17, pp. 13–34). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.17.07bes

Birdsong, D. (2007). Nativelike pronunciation among late learners of French as a second language. In O.-S. Bohn & M. J. Munro (Eds.), *Language Learning & Language Teaching* (Vol. 17, pp. 99–116). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.17.12bir

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer (Version 6.0.37) [Computer program and manual]*. http://www.praat.org/

Bohn, O.-S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition*, 131–158.

Bond, Z. S., & Fokes, J. (1991). Perception of English voicing by native and nonnative adults. *Studies in Second Language Acquisition*, *13*(4), 471–492. https://doi.org/10.1017/S0272263100010299

Bradlow, A. R. (2008a). Training non-native language sound patterns. *Phonology of Second Language Acquisition*, *36*, 287–308.

Bradlow, A. R. (2008b). Training non-native language sound patterns: Lessons from training Japanese adults on the English. *Phonology of Second Language Acquisition*, *36*, 287–308.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*(5), 977–985. https://doi.org/10.3758/BF03206911

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299–2310. https://doi.org/10.1121/1.418276

Caramazza, A., & Yeni-Komshian, G. H. (1974). Voice onset time in two French dialects. *Journal of Phonetics*, *2*(3), 239–245. https://doi.org/10.1016/S0095-4470(19)31274-4

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *The Journal of the Acoustical Society of America*, *54*(2), 421–428. https://doi.org/10.1121/1.1913594

Casillas, J. V. (2018). *lingstuff: Tools and gems for linguistics related research [Computer software manual]* (R package version 0.1.1) [Computer software]. https://github.com/jvcasill/lingStuff

Casillas, J. V. (2020a). The longitudinal development of fine-phonetic detail: Stop production in a domestic immersion program. *Language Learning*, *70*(3), 768–806.

Casillas, J. V. (2020b). Phonetic Category Formation is Perceptually Driven During the Early Stages of Adult L2 Development. *Language and Speech*, *63*(3), 550–581. https://doi.org/10.1177/0023830919866225

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

Catford, J. C., & Pisoni, D. B. (1970). Auditory vs. Articulatory training in exotic sounds. *The Modern Language Journal*, *54*(7), 477–481.

Dalby, J., & Kewley-Port, D. (1999). Explicit pronunciation training using automatic speech recognition technology. *CALICO Journal*, 425–445.

Dalby, J., Kewley-Port, D., & Sillings, R. (1998). Language-specific pronunciation training using the HearSay system. *STiLL-Speech Technology in Language Learning*.

Danan, M. (1992). Reversed subtitling and dual coding theory: New directions for foreign language instruction. *Language Learning*, *42*(4), 497–527. https://doi.org/10.1111/j.1467-1770.1992.tb01042.x

Davidson, L. (2016). Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics*, *54*, 35–50. https://doi.org/10.1016/j.wocn.2015.09.003

Deng, J. (2009). *English pronunciation research: The neglected orphan of second language acquisition studies?* Prairie Metropolis Centre.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 1–16. https://doi.org/10.1017/S0272263197001010

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, *42*(4), 476. https://doi.org/10.1017/S026144480800551X

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, *48*(3), 393–410. https://doi.org/10.1111/0023-8333.00047

Dubreil, S. (2011). Rebels with a cause: (Re)defining identities and culture in contemporary French cinema. *L2 Journal*, *3*, 27.

Elliott, A. R. (1995). Foreign language phonology: Field independence, attitude, and the success of formal instruction in Spanish pronunciation. *The Modern Language Journal*, *79*(4), 530–542. https://doi.org/10.1111/j.1540-4781.1995.tb05456.x

Evans, B. G., & Alshangiti, W. (2018). The perception and production of British English vowels and consonants by Arabic learners of English. *Journal of Phonetics*, *68*, 15–31.

Fischer-Jørgensen, E. (1972). “p t k” et “b d g” français en position intervocalique accentuée. In *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre* (pp. 143–200). De Gruyter Mouton.

Flege, J. (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. https://doi.org/10.1016/S0095-4470(19)30537-6

Flege, J. (1991). Perception and production: The relevance of phonetic input to L2 phonological learning. In T. Huebner & C. A. Ferguson (Eds.), *Crosscurrents in second language acquisition and linguistic theories* (Vol. 2, pp. 249–289). John Benjamins Publishing Company.

Flege, J. (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). York Press.

Flege, J. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development: Papers in honor of Henning Wode*. Wissenschaftlicher Verlag Trier.

Flege, J. (2007). Language contact in bilingualism: Phonetic system interactions. *Laboratory Phonology, 9,* 353–381.

Flege, J. (2009). Give input a chance. In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA.* Multilingual Matters.

Flege, J. (2018). It's input that matters most, not age. *Bilingualism: Language and Cognition, 21*(5), 919–920. https://doi.org/10.1017/S136672891800010X

Flege, J. E. (1982). Laryngeal timing and phonation onset in utterance-initial English stops. *Journal of Phonetics, 10*(2), 177–192. https://doi.org/10.1016/S0095-4470(19)30956-8

Flege, J. E., & Bohn, O.-S. (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., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25*(4), 437–470. https://doi.org/10.1006/jpho.1997.0052

Flege, J. E., & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica, 43*(4), 155–171. https://doi.org/10.1159/000261768

Flege, J. E., MacKay, I. R., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America, 106*(5), 2973–2987.

Flege, J. E., Takagi, N., & Mann, V. (1995). Japanese adults can learn to produce English/I/and/l/accurately. *Language and Speech, 38*(1), 25–55.

Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of/ɹ/and/l. *The Journal of the Acoustical Society of America, 99*(2), 1161–1173.

Flege, J., & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication*, *6*(3), 185–202. https://doi.org/10.1016/0167-6393(87)90025-2

Flege, J., & Eefting, W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *The Journal of the Acoustical Society of America*, *83*(2), 729–740. https://doi.org/10.1121/1.396115

Flege, J., Frieda, E. M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, *25*(2), 169–186. https://doi.org/10.1006/jpho.1996.0040

Flege, J., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *The Journal of the Acoustical Society of America*, *76*(3), 708–721. https://doi.org/10.1121/1.391257

Flege, J., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, 26. https://doi.org/10.1017/S0272263101004041

Flege, J., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, *97*(5), 3125–3134. https://doi.org/10.1121/1.413041

Flege, J., & Schmidt, A. M. (1995). Native speakers of Spanish show rate-dependent processing of English stop consonants. *Phonetica*, *52*(2), 90–111.

Fougeron, C., & Smith, C. L. (1999). *French. Handbook of the International Phonetic Association*. Cambridge University Press Cambridge.

Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition*, *26*(2), 275–301. https://doi.org/10.1017/S0272263104262064

Gabriel, C., Kupisch, T., & Seoudy, J. (2016). VOT in French as a foreign language: A production and perception study with mono-and multilingual learners (German/Mandarin-Chinese). *SHS Web of Conferences*, *27*, 09002.

Gelb, D. (2016, September 2). *Chef's Table: France* (No. 1). Netflix.

Giannakopoulou, A., Brown, H., Clayards, M., & Wonnacott, E. (2017). High or low? Comparing high and low-variability phonetic training in adult and child second language learners. *PeerJ*, *5*, e3209. https://doi.org/10.7717/peerj.3209

González López, V., & Counselman, D. (2013). L2 acquisition and category formation of Spanish voiceless stops by monolingual English novice learners. In J. Cabrelli Amaro, G. Lord, A. de Prada Perez, & J. E. Aaron (Eds.), *Selected Proceedings of the 16th Hispanic Linguistics Symposium* (pp. 118–127). Cascadilla Proceedings Project.

Gonzalez-Bueno, M. (1997). Voice-onset-time in the perception of foreign accent by native listeners of Spanish. *International Review of Applied Linguistics in Language Teaching*, *34*(4), 251–262. https://doi.org/10.1515/iral.1997.35.4.251

Gottfried, T. L. (1984). Effects of consonant context on the perception of French vowels. *Journal of Phonetics*, *12*(2), 91–114. https://doi.org/10.1016/S0095-4470(19)30858-7

Hao, Y.-C., & de Jong, K. (2016). Imitation of second language sounds in relation to L2 perception and production. *Journal of Phonetics*, *54*, 151–168. https://doi.org/10.1016/j.wocn.2015.10.003

Hardison, D. M. (1999). Bimodal speech perception by native and nonnative speakers of English: Factors influencing the McGurk effect. *Language Learning*, *49*, 213–283. https://doi.org/10.1111/j.1467-1770.1996.tb00640.x

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics; New York*, *24*(4), 495–522. http://dx.doi.org/10.1017/S0142716403000250

Hardison, D. M. (2005). Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment. *Applied Psycholinguistics; New York*, *26*(4), 579–596. http://dx.doi.org/10.1017/S0142716405050319

Hazan, V. L., & Boulakia, G. (1993). Perception and production of a voicing contrast by French-English bilinguals. *Language and Speech*, *36*(1), 17–38. https://doi.org/10.1177/002383099303600102

Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., & Chung, H. (2006). The use of visual cues in the perception of non-native consonant contrasts. *The Journal of the Acoustical Society of America*, *119*(3), 1740–1751. https://doi.org/10.1121/1.2166611

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, *47*(3), 360–378. https://doi.org/10.1016/j.specom.2005.04.007

Herron, C., Cole, S. P., Corrie, C., & Dubreil, S. (1999). The effectiveness of a video-based curriculum in teaching culture. *The Modern Language Journal*, *83*(4), 518–533. https://doi.org/10.1111/0026-7902.00038

Herron, C., Dubreil, S., Cole, S. P., & Corrie, C. (2000). Using instructional video to teach culture to beginning foreign language students. *CALICO Journal*, *17*(3), 36.

Hirata, Y., & Kelly, S. D. (2010). Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language and Hearing Research (Online); Rockville*, *53*(2), 298–310. http://dx.doi.org/10.1044/1092-4388(2009/08-0243)

Hombert, J.-M. (1976). The effect of aspiration on the fundamental frequency of the following vowel. *Annual Meeting of the Berkeley Linguistics Society*, *2*, 212–219.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America*, *25*(1), 105–113.

Inceoglu, S. (2016). Effects of perceptual training on second language vowel perception and production. *Applied Psycholinguistics*, *37*(5), 1175–1199. https://doi.org/10.1017/S0142716415000533

Inceoglu, S. (2019). Exploring the perception and production of L2 French vowels: The role of phonological memory. *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, 147–157.

Iverson, G. K., & Salmons, J. C. (2003). Laryngeal enhancement in early Germanic. *Phonology*, *20*(1), 43–74. https://doi.org/10.1017/S0952675703004469

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47–B57.

Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/ contrast by francophones. *Perception & Psychophysics*, *40*(4), 205–215. https://doi.org/10.3758/BF03211500

Jiang, N. (2013). *Conducting reaction time research in second language studies*. Routledge.

Kaiser, M. (2011). New approaches to exploiting film in the foreign language classroom. *L2 Journal*, *3*(2). https://doi.org/10.5070/L23210005

Kaiser, M., & Shibahara, C. (2014). Film as source material in advanced foreign language classes. *L2 Journal*, *6*, 14.

Kazanina, N., Phillips, C., & Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences*, *103*(30), 11381–11386. https://doi.org/10.1073/pnas.0604821103

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, *60*(2), 286–319. https://doi.org/10.2307/413642

Kewley-Port, D., & Watson, C. S. (1994). Formant-frequency discrimination for isolated English vowels. *The Journal of the Acoustical Society of America*, *95*(1), 485–496.

Kingston, J. (2003). Learning foreign vowels. *Language and Speech*, *46*(2–3), 295–348. https://doi.org/10.1177/00238309030460020201

Kingston, J., & Diehl, R. L. (1994). Phonetic knowledge. *Language*, *70*(3), 419–454. https://doi.org/10.1353/lan.1994.0023

Kirby, J. P., & Ladd, D. R. (2015). Stop voicing and F0 perturbations: Evidence from French and Italian. *ICPhS*.

Kleber, F., Harrington, J., & Reubold, U. (2012). The relationship between the perception and production of coarticulation during a sound change in progress. *Language and Speech*, *55*(3), 383–405. https://doi.org/10.1177/0023830911422194

Laguitton, V. (1997). *Indices acoustiques et perception de la parole: Nature et latéralisation hémispherique des processus de traitement (études comportementales et électrophysiologiques)* [PhD Thesis]. Rennes 2.

Lavan, N., Knight, S., Hazan, V., & McGettigan, C. (2019). The effects of high variability training on voice identity learning. *Cognition*, *193*, 104026. https://doi.org/10.1016/j.cognition.2019.104026

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*(3), 369–377. https://doi.org/10.2307/3588485

Levis, J. M., & Grant, L. (2003). Integrating pronunciation into ESL/EFL classrooms. *TESOL Journal*, *12*(2), 13–19.

Levy, E. S. (2004). *Effects of language experience and consonantal context on perception of French front rounded vowels by adult American English learners of French* [Ph.D., City University of New York]. https://search.proquest.com/docview/305202631/abstract/6D42232FFC1749E9PQ/1

Levy, E. S. (2009a). Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. *The Journal of the Acoustical Society of America*, *125*(2), 1138–1152. https://doi.org/10.1121/1.3050256

Levy, E. S. (2009b). On the assimilation-discrimination relationship in American English adults' French vowel learning. *The Journal of the Acoustical Society of America*, *126*(5), 2670–2682. https://doi.org/10.1121/1.3224715

Levy, E. S., & Law, F. (2008). Production of Parisian French front rounded vowels by second-language learners. *The Journal of the Acoustical Society of America*, *123*(5), 3073–3073. https://doi.org/10.1121/1.2932851

Levy, E. S., & Law, F. F. (2010a). Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship. *The Journal of the Acoustical Society of America*, *128*(3), 1290. https://doi.org/10.1121/1.3466879

Levy, E. S., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics*, *36*(1), 141–157. https://doi.org/10.1016/j.wocn.2007.03.001

Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech*, *1*(3), 153–167.

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling/b/versus/p/in trochees. *Language and Speech, 29*(1), 3–11.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Lisker, L., & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech*, *10*(1), 1–28. https://doi.org/10.1177/002383096701000101

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences*, *563*, 563–567.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English/r/and/l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America, 94*(3), 1242–1255. https://doi.org/10.1121/1.408177

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English/r/and/l/. III. Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America*, *96*(4), 2076–2087.

Llanos, F., Dmitrieva, O., Shultz, A., & Francis, A. L. (2013). Auditory enhancement and second language experience in Spanish and English weighting of secondary voicing cues. *The Journal of the Acoustical Society of America*, *134*(3), 2213–2224.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874–886. https://doi.org/10.1121/1.1894649

Lord, G. (2008a). Podcasting communities and second language pronunciation. *Foreign Language Annals*, *41*(2), 364–379. https://doi.org/10.1111/j.1944-9720.2008.tb03297.x

Lord, G. (2008b). Second language acquisition and first language phonological modification. In J. B. de Garavito & E. Valenzuela (Eds.), *Selected proceedings of the 10th hispanic linguistics symposium* (pp. 184–193). Cascadilla Proceedings Project.

Louër, D. (forthcoming). VOT of French in contact with English: A comparative study of bilingual speakers in southwestern Ontario. *Proceedings of LACUS Forum 46*, 10.

Mack, M. (1989). Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals. *Perception & Psychophysics*, *46*(2), 187–200.

MacKay, I. R. A., Flege, J., Piske, T., & Schirru, C. (2001). Category restructuring during second-language speech acquisition. *The Journal of the Acoustical Society of America*, *110*(1), 516–528. https://doi.org/10.1121/1.1377287

McCloy, D. R. (2012). Vowel normalization and plotting with the phonR package. *Technical Reports of the UW Linguistic Phonetics Laboratory*, 8.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748.

Medina, V., Hoonhorst, I., Bogliotti, C., & Serniclaes, W. (2010). Development of voicing perception in French: Comparing adults, adolescents, and children. *Journal of Phonetics*, *38*(4), 493–503. https://doi.org/10.1016/j.wocn.2010.06.002

Mitterer, H., & McQueen, J. M. (2009). Foreign Subtitles Help but Native-Language Subtitles Harm Foreign Speech Perception. *PLOS ONE*, *4*(11), e7785. https://doi.org/10.1371/journal.pone.0007785

Munro, M. J. (1987). *Voicing contrasts in French and English labial stops*.

Nagle, C. L. (2017). A longitudinal study of voice onset time development in L2 Spanish stops. *Applied Linguistics*, *40*(1), 86–107. https://doi.org/10.1093/applin/amx011

Nagle, C. L. (2021). Revisiting perception–production relationships: Exploring a new approach to investigate perception as a time-varying predictor. *Language Learning*, *71*(1), 243–279. https://doi.org/10.1111/lang.12431

Nagle, C. L., & Baese-Berk, M. M. (2021). Advancing the state of the art in L2 speech perception-production research: Revisiting theoretical assumptions and methodological practices. *Studies in Second Language Acquisition*, 1–26. https://doi.org/10.1017/S0272263121000371

Nearey, T. M., & Rochet, B. L. (1994). Effects of place of articulation and vowel context on VOT production and perception for French and English stops. *Journal of the International Phonetic Association*, *24*(01), 1. https://doi.org/10.1017/S0025100300004965

Netelenbos, N., Li, F., & Rosen, N. (2016). Stop consonant production of French immersion students in Western Canada: A study of voice onset time. *International Journal of Bilingualism*, *20*(3), 346–357. https://doi.org/10.1177/1367006914564566

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, *39*(2), 132–142. https://doi.org/10.1016/j.wocn.2010.12.007

Nishi, K., & Kewley-Port, D. (2005). Training Japanese listeners to identify American English vowels. *The Journal of the Acoustical Society of America*, *117*(4), 2401–2401. https://doi.org/10.1121/1.4786064

Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research*, *50*(6), 1496–1509. https://doi.org/10.1044/1092-4388(2007/103)

Nishi, K., & Kewley-Port, D. (2008). Nonnative speech perception training using vowel subsets: Effects of vowels in sets and order of training. *Journal of Speech, Language, and Hearing Research*, *51*(6), 1480–1493. https://doi.org/10.1044/1092-4388(2008/07-0109)

O'Brien, M. G. (2004). Pronunciation Matters. *Die Unterrichtspraxis / Teaching German*, *37*(1), 1–9.

Offerman, H. M. (2014). *The effect of visual feedback on voice onset time productions by L2 learners of Spanish*. Purdue University.

Ohde, R. N. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *The Journal of the Acoustical Society of America*, *75*(1), 224–230.

Olson, D. J. (2014). Phonetics and technology in the classroom: A practical approach to using speech analysis software in second-language pronunciation instruction. *Hispania*, *97*(1), 47–68. JSTOR. https://doi.org/10.1353/hpn.2014.0030

Ortega-Llebaria, M., Faulkner, A., & Hazan, V. (2001). Auditory-visual L2 speech perception: Effects of visual cues and acoustic-phonetic context for Spanish learners of English. *AVSP 2001-International Conference on Auditory-Visual Speech Processing*.

Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception, & Psychophysics*, *72*(8), 2254–2264. https://doi.org/10.3758/BF03196699

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 297–314. https://doi.org/10.1037/0096-1523.8.2.297

Polka, L. (1995). Linguistic influences in adult perception of non-native vowel contrasts. *The Journal of the Acoustical Society of America*, *97*(2), 1286–1296.

Prolific. (2014). https://www.prolific.co

Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *The Journal of the Acoustical Society of America*, *119*(3), 1684–1696. https://doi.org/10.1121/1.2161427

Qualtrics. (2005). *Qualtrics*. https://www.qualtrics.com

R Core Team. (2019). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. https://www.R-project.org/

Reeder, J. T. (1998). English speakers' acquisition of voiceless stops and trills in L2 Spanish. *Texas Papers in Foreign Language Education*, *3*(3), 101–118.

Repp, B. H. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech*, *22*(2), 173–189.

Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, *379*, 410.

Rochet, B. L., & Chen, F. (1992). Acquisition of the French VOT contrasts by adult speakers of Mandarin Chinese. *Second International Conference on Spoken Language Processing*.

Rochet, B. L., Nearey, T. M., & Munro, M. J. (1987). Effects of voicing, place, and vowel context on VOT for French and English stops. *The Journal of the Acoustical Society of America*, *81*(S1), S65–S65.

Ross, J., Lilley, K., Clopper, C., Pardo, J., & Levi, S. (forthcoming). *Effects of dialect-specific features and familiarity on cross-dialect phoneticconvergence*.

Saito, K., & van Poeteren, K. (2018). The perception–production link revisited: The case of Japanese learners' English/ɹ/performance. *International Journal of Applied Linguistics*, *28*(1), 3–17.

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, *52*, 183–204. https://doi.org/10.1016/j.wocn.2015.07.003

Schoonmaker-Gates, E. (2015). On voice-onset time as a cue to foreign accent in spanish: Native and nonnative perceptions. *Hispania*, *98*(4), 779–791. JSTOR. https://doi.org/10.1353/hpn.2015.0110

Schuhmann, K. S., & Huffman, M. K. (2019). Development of L2 Spanish VOT before and after a brief pronunciation training session. *Journal of Second Language Pronunciation*, *5*(3), 402–434. https://doi.org/10.1075/jslp.18018.sch

Segalowitz, N., Freed, B., Collentine, J., Lafford, B., Lazar, N., & Díaz-Campos, M. (2004). A comparison of Spanish second language acquisition in two different learning contexts: Study abroad and the domestic classroom. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *10*, 1–18.

Serniclaes, W. (1987). Etude expérimentale de la perception du trait de voisement des occlusives du français [Experimental study of the perception of the voicing feature in French stop consonants]. *Unpublished Doctoral Dissertation. Université Libre de Bruxelles*.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(3), 243–261. https://doi.org/10.1017/S0142716400001417

Shen, C.-N. J. (2019). *Effectiveness of Audiovisual Training on Non-Native English Speech Production and Perception*. https://scholarsbank.uoregon.edu/xmlui/handle/1794/25058

Sherman, J. (2006). *Using authentic video in the language classroom (Repr.)*. Cambridge: Cambridge Univ. Press.

Stevens, J. J. (2002). *The acquisition of L2 Spanish pronunciation in a study abroad context.* University of Southern California.

Stevens, J. J. (2011). Vowel duration in second language Spanish vowels: Study abroad versus at-home learners. *Journal of Second Language Acquisition and Teaching*, *18*, 77–103.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America*, *111*(4), 1872–1891. https://doi.org/10.1121/1.1458026

Stevens, K. N., & Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *The Journal of the Acoustical Society of America*, *55*(3), 653–659.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, *36*(2), 131–145.

Strange, W., Levy, E. S., & Law, F. F. (2009). Cross-language categorization of French and German vowels by naïve American listeners. *The Journal of the Acoustical Society of America*, *126*(3), 1461–1476. https://doi.org/10.1121/1.3179666

Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., & Nishi, K. (2007). Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *The Journal of the Acoustical Society of America*, *122*(2), 1111–1129. https://doi.org/10.1121/1.2749716

Sturm, J. L. (2012). Hey Good Lookin', Whatcha Got Cookin'? "Tout à côté cuisine" in the Intermediate French Classroom. *The French Review*, *86*(2), 322–332.

Sturm, J. L. (2019). Current approaches to pronunciation instruction: A longitudinal case study in French. *Foreign Language Annals*, *52*(1), 32–44. https://doi.org/10.1111/flan.12376

Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *Calico Journal*, *28*(3), 744–765.

Thomson, R. I. (2012). Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Language Learning*, *62*(4), 1231–1258.

Thomson, R. I., Derwing, T. M., Levis, J., Le, H., Lucic, I., Simpson, E., & Vo, S. (2016). Is phonemic training using nonsense or real words more effective. *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference*, 88–97.

Tognozzi, E. (2010). Teaching and evaluating language and culture through film. *Italica*, *87*(1), 24.

Tranel, B., & Bernard, T. (1987). *The Sounds of French: An Introduction*. Cambridge University Press.

Traunmüller, H., & Öhrström, N. (2007). Audiovisual perception of openness and lip rounding in front vowels. *Journal of Phonetics*, *35*(2), 244–258. https://doi.org/10.1016/j.wocn.2006.03.002

Tyler, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica*, *71*(1), 4–21. https://doi.org/10.1159/000356237

Vanderplank, R. (2016). 'Effects of' and 'effects with' captions: How exactly does watching a TV programme with same-language subtitles make a difference to language learners? *Language Teaching*, *49*(02), 235–250. https://doi.org/10.1017/S0261444813000207

Wang, Y., Behne, D. M., & Jiang, H. (2008). Linguistic experience and audio-visual perception of non-native fricatives. *The Journal of the Acoustical Society of America*, *124*(3), 1716–1726. https://doi.org/10.1121/1.2956483

Wang, Y., Behne, D. M., & Jiang, H. (2009). Influence of native language phonetic system on audio-visual speech perception. *Journal of Phonetics*, *37*(3), 344–356. https://doi.org/10.1016/j.wocn.2009.04.002

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, *113*(2), 1033–1043. https://doi.org/10.1121/1.1531176

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, *106*(6), 3649–3658. https://doi.org/10.1121/1.428217

Weinberger, S. H., & Kunath, S. A. (2011). The Speech Accent Archive: Towards a typology of English accents. In *Corpus-based Studies in Language Use, Language Learning, and Language Documentation* (pp. 265–281). Brill Rodopi.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, *37*(1), 35–44. https://doi.org/10.3758/BF03207136

Westbury, J. R. (1979). Aspects of the temporal control of voicing in consonant clusters in English. *Texas Linguistic Forum Austin, Tex*, *14*, 1–304.

Wiener, S., Chan, M. K. M., & Ito, K. (2020). Do explicit instruction and high variability phonetic training improve nonnative speakers' Mandarin tone productions? *The Modern Language Journal*, *104*(1), 152–168. https://doi.org/10.1111/modl.12619

Wisniewska, N., & Mora, J. C. (2020). Can captioned video benefit second language pronunciation? *Studies in Second Language Acquisition*, *42*(3), 599–624. https://doi.org/10.1017/S0272263120000029

Wong, J. (2012). Training the perception and production of English /e/ and /æ/of Cantonese ESL learners: A comparison of low vs. High variability phonetic training. *Proceedings of the 14th Australasian International Conference on Speech Science and Technology. Sydney, Australia*, 3–6.

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, *79*(7), 2064–2072. https://doi.org/10.3758/s13414-017-1361-2

Yang, Y., Chen, X., & Xiao, Q. (2020). Cross-linguistic similarity in L2 speech learning: Evidence from the acquisition of Russian stop contrasts by Mandarin speakers. *Second Language Research*, 026765831990091. https://doi.org/10.1177/0267658319900919

Zampini, M. L. (1998). The relationship between the production and perception of Spanish stops. *Texas Papers in Foreign Language Education*, 85–100.

Zhang, L. (2011). Teaching Chinese cultural perspectives through film. *L2 Journal*, *3*(2). https://doi.org/10.5070/L23210004