Processing Fluency, Perceptual Adaptation, and Language Attitudes: Does Adaptation Improve Comprehension Ease and Attitudes Toward Speakers with Non-Native Accents?

by

David C. Ogden

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Linguistics)
in the University of Michigan
2020

Doctoral Committee:

Professor Patrice Speeter Beddor, Chair
Associate Professor Jonathan Brennan
Professor Robin Queen
Professor Denise Sekaquaptewa

David C. Ogden

ogdend@umich.edu

ORCID iD: 0000-0003-3880-594X

ACKNOWLEDGMENTS

My heartfelt and undying gratitude goes to my adviser Pam Beddor, who supported and guided me through many difficulties in work and life with skill, compassion, and superhuman generosity. More than anything else, knowing that Pam believed in the value of this research and my ability to do it kept me going. She is an exceptional scientist, skilled editor, and dedicated scholar, and I am grateful to call her a trusted mentor and friend.

From the beginning, this dissertation has benefited greatly thanks to the advice and input of my committee members: Jonathan Brennan, Robin Queen, and Denise Sekaquaptewa. My sincere thanks go to each of them for sharing their time and expertise. I also had the privilege of working with Jon and Robin in graduate coursework, department service, and undergraduate teaching, and I thank them for everything they taught me by example and advice.

Many other faculty were critical in my development as a researcher. Carmel O'Shannessy nurtured my interests in bilingualism and language acquisition and was an excellent adviser for my qualifying research project. I also had the pleasure and honor to work with her documenting the phonetics and phonology of Light Warlpiri. Nick Ellis gave me opportunities and guidance to gain research and technical skills that served me every day of my graduate studies. It was Nick who guided me to my first academic conferences and connected me with preeminent scholars in the field of language acquisition. I deeply enjoyed our time working together and thank him for his kindness and confidence. I also benefited greatly from research guidance, teaching support, and life advice from Andries Coetzee, Jelena Krivokapic, Deborah Keller-Cohen, Marlyse

Baptista, Acrisio Pires, Diane Larsen-Freeman, Julie Boland, Rick Lewis, Anne Curzan, Arnold Ho, Kevin McGowan, Tessa Bent, Melissa Baese-Berk, and Kristin Van Engen. Many thanks to these generous scholars for their support and advice.

Thank you to Ann Bradlow and the Northwestern University Linguistics Labs for sharing the non-native speaker recordings I consulted in creating my experimental stimuli, and to the Northwestern Phonatics discussion group who provided valuable feedback.

I was fortunate to share this research with three discussion groups at the University of Michigan: Psycholinguistics; Phonetics and Phonology; and Sociolinguistics, Language Contact, and Discourse Analysis. Thanks to the members of these groups for their camaraderie and constructive criticism.

Jennifer Nguyen, Sandie Petee, and Talisha Reviere-Winston in the administrative office helped me through so many money, mail, and organizational issues and I'm grateful for all their work and kindness. Thanks also to Jen for her mentorship in navigating degree requirements, tackling the dissertation, and exploring careers.

My list of friends to thank is long, and our lives overlapped in so many ways that it would take pages to acknowledge all of their help. Some were my lab mates and gave me great research ideas and technical help. Many helped me with pilot recordings and data. All of them were with me through hopeless days and crises as well as successes and celebrations. My special thanks to Hayley Heaton, who gave me her guest room for many, many trips back to Ann Arbor to collect data, for her generosity and friendship. Dominique Bouavichith, Tamarae Hildebrandt, Justin Craft, and Dan and Alicia Stevers also put me up more than once and helped in many other ways. And for their commiseration, advice, rescues, game nights, parties, dog-sitting, conference company, and all around great friendship, thanks to Batia Snir, Sagan Blue, Kate

Sherwood, Stephen Tyndall, Will Nediger, Marcus Berger, Ian Calloway, Marjorie Herbert, Kailen Shantz, Emily Cibelli, Eleanor Chodroff, Myrna Cintron-Valentin, Jiseung Kim, Harim Kwon, Paige Safyer, Andrew McInnerny, Rachel Weissler, Kelly Wright, Jian Zhu, Will Styler, Yourdanis Sedarous, Carrie Ann Morgan, Emily Sabo, Dominique Canning, Joy Peltier, and Erica Beck.

Finally, my love and gratitude goes to my wonderful family. Thanks to my parents, who gave me every opportunity without thought for themselves, and to my sisters, who supplemented my formal education with their own smarts and moxie. Buddy and Edie were my faithful writing companions and unfailing rays of happiness. And to Ryan, who upended his life to follow my Ph.D. ambition and held me together through it all, my love and gratitude forever.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABSTRACT

Negative attitudes toward non-native speakers of English are well-documented and have adverse impacts on English learners in a variety of settings. Recent research has proposed that (i) disfluent processing of accented speech (i.e., the metacognitive feeling of effort that accompanies perceiving and integrating information) and (ii) stereotypes and categories associated with specific accents both play a role in these negative attitudes. Previous research has also shown that comprehension of accented speech becomes more accurate over time as listeners perceptually adapt to unfamiliar speech patterns. This dissertation tests the hypothesis that adaptation through listening experience leads to less effortful processing and thereby more positive attitudes, and more broadly investigates the relationship among processing fluency, the subjective perception of fluency, and attitudes toward a speaker with a non-native accent.

In two experiments, participants listened to and transcribed sentences recorded by a speaker in either a non-native or native guise. The non-native accent was selected, based on pre-tests, to be difficult to categorize (e.g., as a native speaker of Spanish), in order to minimize attitude stereotypes associated with judgments of ethnicity or nationality and better isolate the influence of processing fluency. In Experiment 1, all participants listened to the sentences in clear audio. In Experiment 2, participants listened to sentences either in clear audio or mixed with speech-shaped noise, which provided an additional manipulation of processing fluency. Objective fluency was assessed using transcription accuracy in both experiments, in addition to pupil dilation during listening and transcription time in Experiment 1. Subjective fluency and attitudes toward the speaker on the dimensions of warmth, competence, and social closeness were measured at intervals using scale-response questions.

The results show that pupil dilation and transcription time decrease with experience with a non-native accent, indicating improved objective processing fluency, though transcription accuracy results unexpectedly did not show evidence for adaptation to the accent. The results also provide no evidence that listeners' subjective perception of fluency or attitudes toward the speaker change as objective fluency improves. Analysis of the relationship between these variables indicates that listeners' perception of effort in comprehending non-native accented speech is more closely related to their attitudes and social evaluations of the speaker, rather than to objective processing effort. This dissertation adds two main contributions to research on the comprehension of and attitudes toward non-native speakers. First, it provides evidence that comprehension of non-native accented speech becomes easier over time using a physiological measure (pupil dilation) in addition to behavioral measures. Second, by comparing objective fluency, subjective fluency, and attitudes in the same task, it shows that the perception of effort does not necessarily reflect objective effort in comprehending accented speech, but instead appears to be associated with listeners' social judgments. For listeners who experience difficulty understanding non-native accented speech, comprehension is likely to get easier in a short period of time as their cognitive systems adapt to unfamiliar speech patterns. The more a listener recognizes this increased ease, the more that ease may lead to improved attitudes toward the speaker.

CHAPTER 1

Introduction

It has been argued that language, including accent, is one of the last domains in which discrimination is widely acceptable (Lippi-Green, 2012). Research on language attitudes and discrimination has predominantly focused on judgments driven by social categories or stereotypes. In this account, listeners use accent to identify a speaker's social group membership (e.g., ethnicity, geographic region, or native language), which then influences their evaluation of (i.e., attitude toward) the speaker. However, research in social psychology has shown that factors other than social categories influence attitudes, including *processing fluency*: the ease or effort involved in processing information. Listeners often experience difficulty comprehending accented speech, and recent research suggests that processing disfluency is a factor in negative attitudes toward a speaker with an accent, apart from social category judgments (Dragojevic et al., 2017; Dragojevic & Giles, 2016).

Non-native accents in particular are commonly associated with negative evaluation and difficult comprehension. To the extent that negative evaluation of a non-native speaker is due to effortful processing, increasing processing fluency may ameliorate that negativity. Indeed, non-native speakers who experience stigma because of their accent frequently feel pressure to change their speech patterns to be easier to understand (Derwing, 2003; Gluszek & Dovidio, 2010b). While comprehension of non-native accented speech is commonly framed as the responsibility of the speaker, successful communication also depends on the listener. Perceptual adaptation research in phonetics and speech science has shown that listener accuracy in comprehending accented speech improves measurably even with only brief experience (Adank & Janse, 2010; Baese-Berk et al., 2013; Bradlow & Bent, 2008; Clarke & Garrett, 2004; Sidaras et al., 2009). This research suggests that listening experience may also be a means of increasing processing fluency and thereby improving attitudes toward non-native speakers.

This dissertation brings together these research strands to test the hypothesis that perceptual adaptation to non-native accented speech improves both processing fluency and attitudes toward the speaker. These predictions were tested in two experiments combining methodologies from perceptual adaptation, processing fluency, social perception, and language attitudes research. In addition, the methodologies developed here to test the main hypothesis are used to replicate and extend previous research. In the remainder of this chapter, I review that research and present the specific novel hypotheses tested here.

1.1 Social Evaluation of Accented Speakers

Social perception research has found that evaluations of a person can be grouped into two broad domains: warmth and competence. Warmth encompasses "traits that are related to perceived intent, including friendliness, helpfulness, sincerity, trustworthiness, and morality" while competence includes "traits that are related to perceived ability, including intelligence, skill, creativity, and efficacy" (Fiske et al., 2007, p. 77) Language attitudes research has traditionally used a similar division but different labels: solidarity and status (see Dragojevic & Giles, 2016; Heaton & Nygaard, 2011).[1]

Language attitudes researchers consistently find that participants negatively evaluate speakers and accents perceived as non-standard (see Lindemann & Subtirelu, 2013; Purnell et al., 1999). These evaluations have been documented in educational settings, both for adult language learners as well as for instructors who are not native speakers of the language of instruction (Kang et al., 2014; Kang & Rubin, 2009; Rubin, 1992; Subtirelu & Lindemann, 2014), and in professional life, where accent has been a factor in discriminatory hiring practices and workplace behavior (Gluszek & Dovidio, 2010a; Lippi-Green, 2012, pp. 152–156). Though researchers have only recently begun to systematically examine the subjective experience of accented speakers, recent work reports that non-native speakers experience stigma and discrimination because of their accent (Gluszek & Dovidio, 2010b).

Accents are strong cues to social identity and convey information about who we are and where we come from. As a result, many accents are associated with broadly-shared cultural stereotypes and social status. Social categories are undoubtedly a powerful driver of language attitudes towards non-native speakers: listeners commonly use speech patterns to infer a speaker's nationality, ethnicity, or native language, with varying degrees of accuracy and precision. These inferences bring with them a constellation of stereotypes and social judgments that influence, whether implicitly or explicitly, attitudes toward the speaker. Listeners also have stereotypes about non-native speakers as a social category. Non-native English speakers are

[1] Throughout this dissertation, social perception is framed in terms of warmth and competence. Where previous research has used different terms or specific traits, their relationship to warmth and competence is noted.

often stereotyped as less loyal (a warmth trait) and less competent than native speakers (see Gluszek & Dovidio, 2010a) and more difficult to comprehend (Lindemann, 2005). Native speakers both expect and report experiencing greater communication difficulty in interacting with non-native speakers (Derwing et al., 2002; Gluszek & Dovidio, 2010b).

1.2 Processing Fluency and Language Attitudes

Processing fluency is the metacognitive feeling of ease or effort that accompanies the cognitive processes recruited in perceiving and integrating information. A large body of research has shown that fluency operates in a wide variety of cognitive domains, from basic processes like sensory perception to higher-order processes like imagination. Likewise, the experience of fluency provides information used in a variety of judgments, including truth, confidence, and liking (Alter & Oppenheimer, 2009; Schwarz, 2004). This section reviews the hypothesized links between processing fluency and social judgments and the effects of fluency in linguistic processing on evaluations of a speaker.

1.2.1 Fluency and Judgments of Liking and Preference

Processing fluency has been shown to influence liking in a variety of domains. Simple visual stimuli are preferred when they are are easier to process, whether because of previous exposure, longer presentation, enhanced visual contrast, or subliminal priming with matching visual contours or semantically related words (see Alter & Oppenheimer, 2009; Schwarz, 2004 for reviews). Winkielman, Halberstadt, Fazendeiro, and Catty (2006) showed that the preference for prototypical faces extends to abstract visual patterns, indicating that the greater ease with which people process prototypical stimuli influences judgments of attractiveness. In the linguistic domain, sentence-final words that are preceded by a semantically predictive context are judged to be more pleasant than words preceded by a non-predictive context (Whittlesea, 1993).

The links between processing fluency and liking are complex and can be context-dependent. Reasoning and inference about the source and meaning of fluency are largely governed by so-called *naïve theories*, which are learned associations or correlations between fluency and other phenomena. For example, because familiarity (previous experience) results in more fluent processing, items that are easy to process are judged to be familiar. Naïve theories are multifarious, reflecting the wide variety of contexts, cognitive processes, and judgments in which processing fluency is implicated. Because of this, naïve theories can be subject to misattribution or misinterpretation. For example, Whittlesea (1993) used visual masking or semantic context to manipulate the processing fluency of words presented on a screen. Participants were more likely to judge that they had seen a novel word before when processing was actually facilitated by another factor. Task design and explicit instruction can also affect the inferences attached to fluency. When a participant is asked to rate the “brightness” of a visual stimulus, greater ease increases the perception of brightness; when asked to rate “darkness,” greater ease decreases the perception of brightness (Mandler et al., 1987). Briñol, Petty, and Tormala (2006) showed that support for a proposed exam policy was reversed depending on whether students were told that easy processing is a characteristic of intelligent or unintelligent people.

Psychologists have debated whether fluent processing can influence judgments of both liking and disliking, depending on context, in the same way it affects judgments of, for example, brightness and darkness. In a review of this debate, Winkielman, Schwarz, Fazendeiro, and

Reber (2003) surveyed a variety of evidence showing that ease is fundamentally positive in judgments of preference. First, there are general naïve theories about the meaning of processing fluency that contribute to this connection. Ease is a basic cue to judgments of familiarity, prototypicality, and cognitive progress, and all of these phenomena appear to be closely associated with positive valence, perhaps because of their respective evolutionary utility in detecting danger, identifying desirable mates, and comprehending sensory experience. Second, the positive affect elicited by fluent processing provides a more direct and less context-sensitive association between ease and positive evaluation. Winkielman and Cacioppo (2001) measured the response of facial muscles that index affective response and found that fluent processing is associated with positive affect. Additionally, regardless of whether participants are asked to report "liking" or "disliking", greater fluency is associated with more liking and less disliking (Reber et al., 1998; Winkielman et al., 2003).

In examining the links between fluency and evaluation, previous research has distinguished between "objective" and "subjective" fluency. As Winkielman et al. (2003) explain, objective fluency refers to "a mental process characterized by high speed, low resource demands, high accuracy, or other indicators of efficient processing," while subjective fluency is the experience of that ease in conscious awareness. Experimentally, objective fluency is operationalized using behavioral and physiological measures of cognitive resource demands or, more commonly, by manipulating such demands directly (e.g., increasing or decreasing the presentation duration of a stimulus); subjective fluency is measured using participants' self-report. The distinction is motivated by the observation that participants' own report of ease or difficulty is sometimes dissociated from behavioral or physiological measures of effort. For example, objective measurement and subjective experience may reflect fluency during different stages of processing. Reber, Wurtz, and Zimmerman (2004) found that low visual contrast had an effect on word detection responses (i.e., “Is there a word on the screen?”) but no effect on word identification responses (i.e., “What word is on the screen?”). Conversely, a difficult-to-read font had only a marginal effect on word detection but a strong effect on word identification. The results indicate that different processing stages (in this case, detection and identification) are affected by different manipulations of objective processing fluency. Subjective fluency ratings, however, were affected by both visual contrast and font, indicating that subjective fluency incorporates experience from both processing stages.

Objective and subjective fluency also affect evaluation in different ways. Objective fluency affects judgment through automatic processes, such as positive affect, while subjective fluency is used as information in theory-driven reasoning and depends on perceivers’ inferences about the source and meaning of fluency (Forster et al., 2016; Winkielman et al., 2003). As mentioned previously, reasoning based on naïve theories is susceptible to misattribution or misidentification based on the perceiver’s beliefs, and naïve theories may influence the perception of fluency itself, leading to subjective fluency that is not a direct reflection of objective fluency.

The relationship between objective and subjective fluency and their effects on evaluation are somewhat unclear, as few studies have examined them together. Forster, Leder, and Ansorge (2013) measured subjective fluency ratings and liking for line drawings of everyday objects. When objective fluency was manipulated by a subliminal prime, subjective fluency, but not objective fluency (i.e., whether the prime was congruent or incongruent), was positively associated with liking evaluation. When objective fluency was manipulated by presentation duration, both subjective and objective fluency (i.e., length of stimulus presentation) were positively associated with liking evaluation, though the effect of subjective fluency was larger. Subjective fluency was positively affected by objective fluency in both fluency manipulations. In another study, von Helversen, Gendolla, Winkielman, and Schmidt (2008) measured support for a public policy proposal and (in a separate experiment) participants' self-rated assertiveness. The researchers manipulated fluency by asking participants to generate a small (easy condition) or large (difficult condition) number of arguments that supported or opposed a positive evaluation and measured objective fluency using participants' heart rate and blood pressure during the argument-generation task. As expected based on previous research, the easy task increased support or opposition consistent with the arguments generated, while the difficult task increased conclusions that were inconsistent with the arguments generated. The researchers found that participants' reported effort (subjective fluency) was related to their physiological response (objective fluency), and both measures indicated greater effort when participants generated a large number of arguments. A mediation analysis showed that only subjective fluency was directly related to evaluative judgments, though regression analysis showed that subjective fluency did not explain variance in judgment over and above the easy or difficult condition to which the participant was assigned. The causal relationship between subjective fluency and evaluation is difficult to assess in this study, as subjective fluency was measured after the evaluative judgment for all participants.

The research reviewed here indicates that high objective fluency is straightforwardly expected to have a positive influence on evaluation, but subjective fluency may mediate that effect in ways that depend on listeners' potentially complex and context-sensitive naïve theories. It should be noted that while processing fluency provides a listener with information that can affect social evaluation, other sources of information, including association or inference based on identification or categorization, also play a role. A listener's reliance on fluency versus other information can be influenced by a variety of factors. In experimental settings, greater task demands have been shown to increase participants' reliance on fluency when making judgments (Whittlesea, 1993). More generally, both low motivation and happy mood have also been linked to reliance on heuristic strategies, including fluency-based evaluation (Schwarz, 2004). Non-native speech may itself also be a contextual factor that favors heuristic judgment, as listeners appear to process non-native speech less deeply, paying less attention to what the person is saying and more attention to contextual cues and expectations to help interpret the message (see Lev-Ari, 2015).

1.2.2 Fluency and Social Evaluation in the Linguistic Domain

The association between fluent processing and positive affect and the general naïve theories that link ease with familiarity, prototypicality, and cognitive progress are implicated in social perception, including the perception and evaluation of speech and speakers. Language attitudes may also be subject to domain-specific naïve theories. When listeners believe that responsibility for successful communication (the communicative burden) lies with the speaker, perhaps due to context or language ideology (see Lippi-Green, 2012), perceptual disfluency is likely to be interpreted as a lack of warmth or competence on the part of the speaker (Dragojevic & Giles, 2016). Consistent with this prediction, students who read passages of text with longer vocabulary items or greater complexity experienced greater disfluency and rated the author as less intelligent (a competence trait) than did students who read a comparable but simpler passage (Oppenheimer, 2006).

Researchers have only recently begun to directly examine the role of processing fluency in language attitudes. Dragojevic and Giles (2016) examined fluency of perception and language attitudes towards Punjabi English (PE; associated with speakers from India and Pakistan) in two experiments with matched guise designs; a single bidialectal female speaker recorded experimental stimuli in two accents so that speaker differences were not confounded with accent differences. In both experiments, two groups of participants heard a recording of the speaker telling a fictional story, one group hearing the Punjabi guise, and the other hearing an American guise ("Standard American English"; SAE). Each of these groups was further divided into five noise conditions, with white noise mixed with the speech recording, ranging from no noise to noise equal in intensity to the speech. After listening to the story, participants reported their affective response by rating their own feeling of three positive and three negative emotions. They then rated the speaker on traits related to status (corresponding to competence, e.g., "intelligent," "competent") and solidarity (corresponding to warmth, e.g., "friendly," "sociable"). Last, they reported the ease or effortfulness they experienced in comprehending the speaker, which was used as the measure of (subjective) processing fluency. Dragojevic and Giles predicted that both the PE guise and noise would have negative effects on processing fluency, affect, and evaluations of the speaker.

In the first experiment, the participants' task was only to listen, with no test for comprehension, and they were not told anything about the speaker's social group or identity. Participants' solidarity ratings, but not status, were lower for the PE guise. In both guises, increased noise had a negative effect on subjective processing fluency. For the other ratings, however, the modeled effects of noise interacted with the effects of guise, such that noisier audio conditions for participants listening to the PE guise, but not the SAE guise, resulted in more negative ratings for status, solidarity, and affect. Mediation analysis of the effect of noise was performed using Hayes' (2013, cited in Dragojevic & Giles, 2016) PROCESS macro, which models the effects of one variable on another via a third variable. In this case, the model indicated that the effect of noise on status and solidarity ratings was mediated by subjective fluency, i.e., higher intensity noise reduced subjective fluency which in turn resulted in lower status and solidarity ratings. The direct effect of subjective fluency on ratings was significant, as was the effect of subjective fluency mediated by affect.

While increased noise had a negative effect on affect and speaker evaluation in the PE guise, consistent with Dragojevic and Giles's prediction, noise also affected how participants categorized the speaker. At the end of the first experiment, participants were asked where they thought the speaker was from. While participants who heard the SAE guise were equally accurate at categorizing the speaker as from California/U.S. West Coast across noise conditions, those who heard the PE guise were much less accurate at categorizing the speaker as from India in high noise (e.g., "foreigner") than in low noise (e.g., "Indian"). As Dragojevic and Giles acknowledge, this difference in categorization may have influenced the stereotypes listeners had about the PE guise in the different noise conditions, which may have influenced ratings of the speaker independent of the subjective fluency effects of noise.

The absence of an effect of noise on attitude ratings in the SAE condition was inconsistent with Dragojevic and Giles's predictions. They posit that this result was due to participants' familiarity with the SAE accent, which allowed for high noise levels without participants feeling that communication was compromised. However, the result may be consistent with Lippi-Green's (2012) argument that the communicative burden is commonly shifted more to speakers who are not in a listener's own social group, making fluency a basis for judgment of the speaker in the PE guise but not the SAE guise. This may mirror what Pearson (2011) found in Chapter 1, Study 2, of his dissertation research, where disfluency resulted in more negative attitudes only for outgroup members (in his study, Black people), but not for ingroup members (White people). Pearson interpreted this as evidence that processing fluency is relied on more for judgment of outgroup members than for ingroup members in general.

In Dragojevic and Giles's second experiment, the experimenter told participants what accent they were about to hear in order to remove variation in social category judgments. In order to make participants more aware of their comprehension accuracy, and to ensure that noise and the PE guise had the expected negative effect on accuracy, the participants were asked to fill in 12 missing words in a transcript of the story they heard. The missing words were not predictable from context. Participants then completed the same affect, attitude, and subjective fluency rating items as in the first experiment.

In the second experiment, as in the first, increased noise had a negative effect on affect, speaker evaluation, and subjective fluency in the PE guise. Since all participants in this condition were told that the speaker was from India, these results are likely not due to different social category and stereotype judgments in the different noise conditions. Unlike in the first experiment, the same effects of noise were also observed for the SAE guise. These results are inconsistent with the hypothesis that the increased communicative burden placed on outgroup speakers leads to judgments based on disfluency: since participants in this condition were told that the speaker was from California (where the experiment was conducted) and heard an SAE accent, the speaker was almost certainly not perceived as an outgroup member. Instead, these results suggest that subjective difficulty in comprehension (made salient for participants in the second experiment by the comprehension test), or the greater experimental task demands added by the comprehension test itself, induced fluency-based judgments. However, language ideology may still play a role, in that listeners may assume that comprehension will be negatively affected by an accented speaker, leading them to make fluency-based evaluations even when comprehension is unaffected or irrelevant. For speakers with a familiar easy accent, disfluency caused by noise may only affect evaluation when difficult comprehension is noteworthy or conspicuous.

In another study focused on speech processing fluency, Lev-Ari and Keysar (2010) examined the effect of fluency on listeners' attitudes towards the information the speaker was communicating. Participants listened to recordings of native and non-native English speakers reading trivia statements written by the experimenter. Based on independent ratings of accent strength, three of the non-native speakers were classified as mildly accented and another three were classified as heavily accented. To enhance the perception that the speakers were acting as messenger for the experimenter and not sharing their own knowledge, listeners also recorded trivia statements, ostensibly to be used for later participants. In one experiment, participants were not told that accent, language, or social identity were of any interest in the experiment; the stated purpose of the study was to understand "intuition in knowledge assessment" (Lev-Ari & Keysar, 2010, p. 1094). Participants rated information conveyed by native speakers as more true than information conveyed by both mild- and heavy-accented speakers, controlling for whether participants already knew the truthfulness of a trivia statement.

In Lev-Ari & Keysar's second experiment, designed to see if listeners could discount disfluency in their judgments, participants were told that the experimenters were examining how difficulty understanding speech might affect whether the communicated information was believed. Participants also rated the difficulty of understanding each speaker. Participants' truthfulness ratings for statements read by mild-accented speakers were roughly equal to ratings for statements read by native-accented speakers, but ratings were significantly lower for statements read by heavy-accented speakers. In addition, read statements that were rated more difficult to understand were also rated less true. Because the experimenters took steps to convince the participants that the speaker was not representing their own knowledge but statements provided by the experimenter, the authors interpreted the observed effects as suggesting that perceptual fluency may affect the trustworthiness (a warmth trait) of accented speakers, particularly those that are difficult to comprehend, independent of inferences about a speaker's trustworthiness based on stereotypes associated with language background or nationality.

The research summarized in this section measured only the relationship between attitude (i.e., social evaluation) and subjective fluency without measuring the relationship between attitude or subjective fluency and objective measures of fluency. Dragojevich and Giles's (2016) second experiment did include a comprehension test, and accuracy on this task (a measure of objective fluency) was significantly reduced by both noisy audio and the Punjabi English guise, as were subjective ratings of fluency, but the relationship of accuracy with subjective fluency, affect, and speaker evaluation was not analyzed. Other studies of non-native speech processing provide some evidence of a positive correlation between objective and subjective fluency. For example, Munro & Derwing (1995b) found that sentences that participants rate as effortful to comprehend (i.e., intelligibility ratings in the bottom third of a nine-point scale) are associated with longer response times for truth judgments of the sentence, relative to sentences with medium or high ratings.[2] In a different study, Munro and Derwing (1995a) measured intelligibility and transcription accuracy for individual sentences. Using Pearson's *r*, they reported significant correlations (i.e., $p < 0.05$) between the two measures for 15 of 18 participants. However, this result may not be reliable because the reported distributions of both variables appear highly skewed and thus inappropriate for the statistical test used. Overall, it is unclear from the limited research on this relation whether objective fluency in language processing is positively correlated with subjective fluency and attitudes toward the speaker.

1.3 Perceptual Adaptation

The prevalence of negative evaluation of non-native speakers, especially insofar as it is driven by disfluent processing, seems at odds with listeners' robust ability to adapt to and understand highly variable speech patterns. This adaptation happens on a small scale everyday, when listeners adjust to the acoustic patterns produced by the different-sized vocal tracts of different-sized people. For example, although a given vowel category (e.g., /u/) has a characteristic pattern of formant (resonant) frequencies that is determined by the vocal tract configuration (tongue position, lip position, etc.) for that vowel, absolute formant frequencies differ across speakers due in part to their different-sized vocal tracts. Ladefoged and Broadbent (1957) found that listeners identify vowels in synthesized speech based in part on the relationship between the formants of the target vowel and other vowels perceived in preceding speech, and

[2] Language acquisition research commonly uses "intelligibility" to refer to objective comprehension accuracy and "comprehensibility" to refer to listener ratings of ease of comprehension. This terminology is not universal and is sometimes reversed. For clarity, the terms used in the text are consistent with usage in the rest of the dissertation (i.e., "intelligibility" refers to a listener's reported subjective experience of processing fluency), not with the terms used in the research cited.

not solely on absolute formant frequency. In other words, it appears listeners rapidly adapt their vowel space to the formant structure of an individual's vocal tract. Similar adaptation occurs with consonants, for example in differentiating the sounds [s] and [ʃ] for female- and male-sounding speakers, the acoustic characteristics of which are also influenced by vocal tract size (Strand, 1999).

Adaptation has also been demonstrated experimentally when listeners encounter speech in an accent that is new or unusual for them. Clarke and Garrett (2004) asked participants to listen to sentences whose final words were not predictable (e.g., "Ruth must have known about the pie") then indicate via button press whether a visually-displayed word matched the final word. Non-matching words were phonetic neighbors (i.e., identical except for one consonant or vowel) of the target word. The reaction time of participants listening to both Spanish- and Chinese- accented speech decreased over the course of listening to 16 different sentences, and in some cases after as few as two to four sentences, and control conditions indicated that this improvement could not be attributed to practice with the task or general adaptation to difficult-to-understand speech. The researchers argued that this decrease in reaction time showed increased processing speed as a result of adaptation specific to the foreign accent.

Bradlow and Bent (2008) asked participants to transcribe sentences spoken by different speakers with Mandarin- and Slovakian-accented English. When participants transcribed sentences from one speaker, their accuracy improved with increased experience. For speakers with lower baseline comprehension accuracy, statistically significant improvement in comprehension required more experience than for speakers with higher baseline comprehension accuracy, but low accuracy did not preclude adaptation. In a separate experiment, participants who listened to multiple accented speakers showed improvement in comprehending a new talker with the same accent, comparable to participants who had been trained only with that talker. However, practice comprehending one native-language accent (e.g., Mandarin), did not improve comprehension of a different accent (e.g., Slovakian). Sidaras, Alexander, and Nygaard (2009) showed similar results for the transcription of both sentences and isolated words for Spanish-accented English. Later experiments by Baese-Berk, Bradlow, and Wright (2013) provided evidence that practice comprehending multiple foreign accents (Thai, Korean, Hindi, Romanian, and Mandarin) improved comprehension of a speaker with a Slovakian accent, though it was unclear whether this adaptation was effectuated by a shifting of phonetic categories in a pattern specific to non-native English accents or a more general relaxation of categories to include less prototypical patterns.

1.4 Research Questions and Hypotheses

The research described in the previous sections indicates that processing fluency is implicated in judgments related to social evaluation (i.e., attitudes toward speakers), and that subjective disfluency is correlated with negative attitudes toward speakers with an accent that is difficult to comprehend. Comprehending non-native speech often presents such difficulty, which may influence the commonly observed negative evaluations of non-native speakers. However, listener comprehension improves with time and experience through perceptual adaptation, suggesting that adaptation may lead to more fluent processing. (It is possible that improved comprehension is the result of increased listener effort to compensate for the difficulty presented by the accent. This ambiguity is addressed in section 2.2.) Integrating these findings, a preliminary model of language attitude change (Figure 1.1) is proposed in which processing is made more fluent through adaptation (a) and improved fluency leads to more positive attitudes towards the speaker (b). (Other factors that are not the focus of the present research are also incorporated, including the influence of attitudes on adaptation [c] and the effects of factors like social categories and stereotypes that have been shown to influence attitudes [d].)

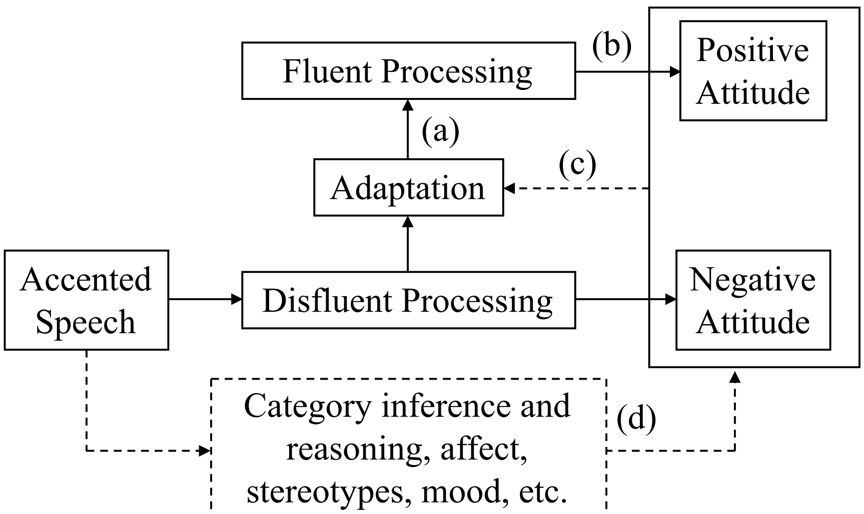

Figure 1.1 A model of the hypothesized effects of adaptation to accented speech on (a) processing fluency and (b) attitudes towards the speaker. Elements of the model not directly tested in the current research include (c) the effects that a listener's attitudes toward accented speech may have on adaptation and (d) other social and individual influences on attitude.

Alternative accounts of the role of processing fluency in attitudes toward non-native speakers are, of course, also compatible with the literature reviewed in this chapter. For example, the research summarized in section 1.1 documents the role of listeners' stereotypes on language attitudes (included in [d] in Figure 1.1), and it is not clear how social information and fluency interact in influencing listeners' evaluation. Social information may influence listeners' perception of processing fluency itself, or strong stereotype-driven attitudes may be insensitive to immediate change based on increased fluency due to adaptation. These alternatives will be considered in chapter 4, in light of this dissertation's findings. At this stage, though, the proposed links between adaptation, fluency, and attitudes in the preliminary model are used to motivate several research questions. First, does perceptual adaptation to accented speech result in increased objective processing fluency (as opposed to improved comprehension outcomes through increased listener effort)? If adaptation does lead to objectively easier processing, do listeners experience that improvement subjectively? Second, is negative evaluation of a non-native speaker related to disfluent processing (objective or subjective), and do evaluations improve over time in tandem with perceptual adaptation? Accordingly, two experiments were designed to evaluate two main hypotheses: (1) processing fluency will increase with exposure to a non-native accent, in terms of both objective (i.e., physiological and behavioral response) and subjective (i.e. self-report) measures; and (2) speaker evaluation will improve with exposure to a non-native accent, in direct relation to the improvement in processing fluency.

1.5 Organization of the Dissertation

In Chapter 2, I motivate and describe the methods, and present predictions and results, for Experiment 1. Chapter 3 presents the methods (many of which overlap with those of Experiment 1), predictions, and results for Experiment 2. Chapter 4 provides a general discussion of findings from both experiments in relation to the research questions and hypotheses described in the previous section, followed by general conclusions and practical implications of the findings and a discussion of how those findings may inform future research in speech perception and language attitudes.

CHAPTER 2

Experiment 1

The hypotheses set out in Chapter 1 were evaluated in two experiments with similar overall designs, tested on separate groups of participants. In Experiment 1, the main hypotheses tested are that (1) processing fluency will increase with exposure to a non-native accent, in terms of both objective (i.e., physiological response via pupillometry and behavioral response via comprehension accuracy and transcription time) and subjective (i.e., self-report) measures; and (2) speaker evaluation will improve with exposure to a non-native accent, in direct relation to the improvement in processing fluency. This chapter first describes the methods and materials used to test these hypotheses (with brief mention, as necessary, of differences between Experiments 1 and 2), followed by predictions, results and statistical analysis, and a brief discussion of the results of Experiment 1. A more comprehensive general discussion of the results of both experiments is presented in Chapter 4.

2.1 Design Overview

In both experiments, participants were randomly assigned to listen to either a non-native (test condition) or native (control condition) accent. They transcribed sentences and evaluated the speaker at multiple points in the experiment in order to measure comprehension accuracy and attitude as well as change over time. In Experiment 1, the sentences were presented in clear audio (i.e., without noise) and processing fluency was measured using the physiological response of pupil dilation, transcription accuracy, transcription time, and self-report. In Experiment 2, participants were randomly assigned to a clear or noisy audio condition (crossed with the non-native or native accent condition) and processing fluency was again measured using transcription accuracy, transcription time, and self-report (but not pupil dilation). Experiment 2 also introduced additional norming procedures for the attitude evaluation items in order to clarify participants' evaluations of the speaker.

2.2 Operationalizing Processing Fluency, Adaptation, and Attitude

Previous research testing adaptation to unfamiliar accents—whether non-native, regional, or artificial—has primarily used accuracy as the dependent variable. In those experiments, participants either transcribed (Baese-Berk et al., 2013; Bradlow & Bent, 2008; Sidaras et al., 2009) or repeated aloud (Adank & Janse, 2010) each utterance they heard and accuracy was scored according to proportion words identified correctly. While these experiments have shown consistent improvement with experience in comprehending an accent, this improvement may be due to increased ease in processing or increased effort on the part of the participant in comprehending. The present research postulates that improvement is due to the former (i.e., increased objective and subjective ease in processing) and that processing disfluency is a driver of negative attitudes toward a speaker with a non-native accent. Experiment 1 addresses the ambiguity in the literature regarding the source of improvement in two ways, both guided by the principle that accurate measurement of any increase or decrease in processing ease is necessary to interpret the relationship between processing fluency and change in speaker evaluation. First, scale-response items were included in the design asking participants to reflect on and rate their subjective experience of ease or difficulty comprehending the speaker's utterances. Second, in order to measure processing fluency objectively and online, i.e. while the participant was listening to each utterance, each participant's pupil diameter was measured using an eye-tracking camera. The pupil dilates in response to cognitive effort (Sirois & Brisson, 2014) and dilation is used here to operationalize processing fluency.

Measurement of pupil dilation as an indicator of objective fluency and adaptation requires participant-specific preparation and calibration of the eye-tracking set-up. Participants are required to sit with their head still, supported by a desk-mounted headrest, for the duration of the experiment—while listening, typing, and answering questions with the computer mouse. These factors limited the number of participants in the experiment and required some design decisions that differed from previous adaptation research. Specifically, although previous adaptation experiments have added noise to recorded stimuli to induce less-than-ceiling performance (to allow for improvement over time), Experiment 1 presented only clear audio because of concern that the combination (and potential interaction) of noise, accent, and the pupillometry testing conditions might overly influence comprehension and speaker evaluation. Experiment 2 excluded pupillometry and instead divided the participants into clear and noisy audio conditions to better assess changes in accuracy with listening experience while also evaluating the interaction of noise and non-native accent.

Attitudes toward the speaker were measured using scale-response items including statements describing the participant's social perception of and social closeness to the speaker (see section 2.4). Social perception was measured by the evaluation of attributes of warmth and competence; section 2.4.2 describes the selection of items to measure these two dimensions. Social closeness items, described in section 2.4.3, were included to evaluate how listeners perceive the potential relationship between themselves and the speaker. This measurement supports the broader aim of this research to provide insight for real-world interaction with non-native speakers. Evaluation of a non-native speaker (tested by social perception items) is predicted to improve with experience, but this improvement may not reduce discrimination if the listener is not receptive to further social contact. Intergroup relations research has also shown that evaluation of outgroup members and social closeness (usually measured by its inverse, social distance) are related (Bastian et al., 2012).

In order to assess perceptual adaptation and attitude change contemporaneously, the scale-response questions were presented at four points over the course of the sentence listening and transcription task. Participants in each experiment heard 100 sentences overall, the first 10 recorded by a native speaker of American English for baseline measurements and the remaining 90 spoken by the target speaker in either a native or non-native guise. Scale-response questions were presented after 10, 20, 60, and 100 sentences, providing baseline, initial, medial, and final measurements, respectively (see Table 2.3 for a more detailed procedure overview). The baseline measurement provides data to compare evaluation of non-native and native accented speech by participants in the test group, and the remaining three are designed to measure if and how responses change with experience listening to an accent.

2.3 Transcription Task Material

2.3.1 Sentences

Perceptual adaptation to a non-native accent depends on some access by the listener to the speaker's intended utterance, despite its unfamiliar realization. However, access alone does not entail adaptation because access can be provided by sentences with semantically familiar and predictable contexts. Measuring adaptation ideally involves testing the recognition of words out of context. Therefore, in both experiments, half of the sentences used had predictable semantic contexts (e.g., “A book tells a story.”), to facilitate adaptation, while the other half did not (e.g., “This is his favorite level.”), to facilitate the measurement of adaptation. These two sets of sentences (see Appendix A) were matched pairwise for the number of words in the sentence and for the frequency and character- and syllable-length of sentence-final words.

The sentences were adapted from those in Bradlow and Alexander (2007), in which pairs of sentences were constructed with the same final word in a predictive or non-predictive context. The predictable final-word sentences were constructed by Bradlow and Alexander using a sentence completion task with both native and non-native speakers to ensure high predictability. Those authors constructed unpredictable sentences with a few frames with limited vocabulary and length. In this study, in order to create sentence pairs that were matched for length in words and more closely matched for lexical variation and syntactic complexity, new unpredictable sentence patterns were constructed using noun phrases, modal verbs, and subordinating clauses based on the Revised Speech Perception In Noise test (see Clarke & Garrett, 2004).

Predictable and unpredictable sentence pairs with identical final words are unsuitable for the current design, which measured change in accuracy over time. Specifically, the repetition of words may interfere with accurate measurement of adaptation because listeners may identify a word without predictive context only because they had previously encountered a phonetically similar realization of the word in a sentence with predictive contextual clues. The present experiments aim to measure general adaptation to the speakers’ speech patterns rather than adaptation to the pronunciation of a single word. To facilitate this, the original final word in each unpredictable sentence (identical to the final word in the corresponding predictable sentence) was replaced with a word matched to the original word in character- and syllable-length, based on the Carnegie Mellon Pronouncing Dictionary (Carnegie Mellon University, 2014), and in frequency, based on the spoken language portion of the British National Corpus (BNC Consortium, 2007).

Sentence presentation order was designed to balance the occurrence of predictable and unpredictable sentences across the experiment. The 100 sentences were divided into ten groups of ten, each with five predictable and unpredictable sentences randomly selected for each participant. Within each group of 10, the order of the sentences was also randomized. This ensured that, (1) at each scale-response rating, every participant had heard an equal number of predictable and unpredictable sentences, and (2) the alternation between sentence types did not follow a pattern that participants could learn. This also balanced any effect that predictive semantic context may have on processing fluency (see Whittlesea, 1993).

2.3.2 Accent for the Non-Native Guise

Because this experiment aims to determine whether negative attitudes arising from processing disfluency improve with adaptation, it is necessary to control for other factors that may affect attitudes negatively but are not amenable to change due to adaptation. The other primary drivers of negative attitudes towards non-native speakers are stereotypes of ethnicities or nationalities that are associated with the accent. An accent that is not strongly associated with a particular place or ethnicity should minimize these effects, providing a better chance of observing attitudes resulting from processing effort.

The following steps were taken to identify an accent with minimal stereotypical associations. Two male speakers were selected from the Wildcat corpus of non-native speaker recordings (Van Engen et al., 2010) who had low intelligibility scores and whose native languages are not commonly stereotyped in the United States: Turkish and Thai. A small convenience sample of non-linguists listened to the two speakers and were asked to report any ethnic, national, or regional identifications they could make. Most participants identified the native Turkish speaker as “Middle Eastern” or “Eastern European,” however the accent of the native Thai speaker was not reliably or specifically identified: nationality responses included locations spread between the Caribbean, Scandinavia, and southeast Asia. Listeners also reported difficulty understanding his speech. Based on these results, the Thai accent was selected as the non-native speech accent to be used in sentence stimuli recorded for this experiment.

Responses to debriefing questions in Experiments 1 and 2 indicated that most participants were unable to identify a nationality or ethnicity associated with the accent. In Experiment 1, one participant suspected that the speaker was from Jamaica or the Caribbean and another reported trying and failing to determine if the speaker’s native language was German. In Experiment 2, two participants concluded that the speaker was from a country in Asia and three others considered Austria, northern Africa, and Boston, but did not seem to make a judgment.

2.3.3 Sentence Recordings

To allow for a matched guise design, in which the same speaker’s voice is presented in both the native and non-native guise, the sentences were recorded by a professional actor imitating the original native Thai speaker from the Wildcat corpus. The actor was not told the purpose of the research or the native language or nationality of the model speaker. He was instructed to imitate the accent, but not the voice, as faithfully as possible. To help his imitation, I drew his attention to several features of the original accent—such as the frequent realization of [ð] as a stop and the non-native-like prosody—in lay terms. I avoided instructing him on specific pronunciation targets to encourage faithfulness to the original rather than faithfulness to categorical instructions.

Sentences in both the native and non-native guise were recorded, with the experimenter present, in a sound-attenuated booth. For the non-native guise, we frequently referenced the original recordings and recorded several takes of each sentence. I selected a single recording of each sentence based on perceived similarity to the original non-native speaker and consistency within the actor's production. No participant guessed that the accent was performed or imitated. In the native guise, his instructions were to sound like a person that a student at the university might meet and consider to be American. The actor had lived and worked in the American Great Lakes region most of his life and his native accent was familiar in the area. He was encouraged to speak naturally rather than targeting any specific style. Most native-accent sentences were recorded only once. In cases where there were more than one recording, I selected the recording that sounded the least like the sentence was read from a list.

2.4 Scale-response Materials

Scale-response items are divided into three categories: intelligibility, social perception, and social closeness. Scale-response ratings occurred four times during each experiment, at timepoints labeled *baseline*, *initial*, *medial*, and *final*. The same intelligibility items were presented in the same order at all four timepoints, with instructions for the participant to consider their "most recent listening experience" when answering. Social perception and social closeness items were the same for the baseline and initial ratings, but different items were used for each of the medial and final ratings. The baseline and initial ratings used the same items to allow comparison between assessment of the baseline speaker, using a general American English accent, and the first assessment of the target speaker in either his native or non-native guise. The items were changed for subsequent ratings to prevent participants from guessing that attitude change was of primary interest in the experiment or reflecting on earlier responses to the same item for the same speaker. Each item listed in the sections below was displayed separately on a screen with a continuous response scale and a slider controlled by the computer mouse. Labels for the response scale differed between the three categories and are noted below.

2.4.1 Intelligibility

In order to assess participants' subjective perception of processing fluency, three statements were designed for responses on an Agree/Disagree response scale:

1. I am able to accurately transcribe what this person says.
2. This person speaks in a way that is easy to understand.
3. I can comprehend this person without difficulty.

These items are intended to measure distinct components of intelligibility in relation to the roles of speaker and listener and the communicative task at hand. Item 1 (the *transcribe* item) asks participants to consider their success at the communicative task in the experiment, without referring to how much effort that success takes. Items 2 and 3 target the participants' feeling of processing fluency: item 2 (the *speaks* item) frames fluency in terms of the speaker's contribution to easy or effortful processing, while item 3 (the *comprehend* item) targets the listener's experience of fluency. The left and right endpoints of the response scale for these items were labelled "Strongly Disagree" and "Strongly Agree," respectively, with no labeled midpoint.

2.4.2 Social Perception

Social perception of the speaker was operationalized using participants' evaluation of the accuracy of short adjective descriptions of the speaker (e.g., "This person is intelligent.") on a scale from "Not at all" to "Very". In order to assess the speaker's potential positive and negative traits of competence and warmth, adjectives were taken from a normed list reported by Bergsieker et al. (2012). In that study, 46 adjectives were identified as referencing either the competence or warmth dimensions and rated for favorability from -2 (very unfavorable) to 2 (very favorable) by undergraduates and a small group of social psychologists at Princeton University. Of the adjectives on Bergsieker and colleagues' list, seven were excluded from the present experiments due to pilot participants' unfamiliarity with the words. Of the 39 remaining adjectives that were used, 15 were related to competence and 24 to warmth.

In order to present different adjective statements for the baseline/initial, medial, and final ratings, these 39 adjectives were split into three sets of 13 adjectives each, 5 for competence and 8 for warmth. Sets were matched as closely as possible on the favorability of the individual items and the mean favorability of the set. Items within the sets were randomly ordered, while across the sets, item order was matched so that participants rated items with the same dimension and similar favorability in the same order. The sets of attitude items are shown in Table 2.1.

Set 1		Set 2		Set 3	
Item	Fav.	Item	Fav.	Item	Fav.
happy-go-lucky	0.75	sensitive	0.66	pleasure-loving	0.57
courteous	1.32	faithful	1.22	passionate	1.02
methodical	0.42	practical	0.87	persistent	0.87
kind	1.46	honest	1.58	generous	1.44
stubborn	-0.78	aggressive	-0.58	reserved	0.04
arrogant	-1.32	quarrelsome	-1.27	rude	-1.46
ignorant	-1.75	lazy	-1.32	stupid	-1.63
ambitious	0.9	industrious	1.09	alert	0.95
efficient	1.12	sophisticated	1.15	intelligent	1.52
cruel	-1.81	deceitful	-1.66	treacherous	-1.59
argumentative	-0.81	quick-tempered	-1.1	boastful	-1.14
shrewd	0.2	naive	-0.91	frivolous	-0.87
humorless	-1.17	conceited	-1.18	revengeful	-1.24

Table 2.1 Adjectives used for attitude evaluation with favorability ratings from Bergsieker et al. (2012). Gray shading indicates a competence attribute; no shading indicates a warmth attribute.

The left and right endpoints of the response scale for the social perception items were labelled “Not at all” and “Very,” respectively. The midpoint of the scale was labelled “Somewhat,” though the precise center of the scale was not marked. Participants were instructed to respond to each statement even if they did not feel they had enough information to evaluate the speaker, but still indicating their best judgment or intuition. In pilot testing, where the midpoint was not labelled, participants sometimes marked the midpoint with the intent of indicating that they had no judgment for that adjective. The “Somewhat” midpoint label was therefore added to reinforce that no non-response was available.

2.4.3 Social Closeness

Social closeness was evaluated in three domains: working, socializing, and living. For each domain, Agree/Disagree statements were developed with input from pilot participants and other students. Three statements were chosen for each domain for use at the baseline/initial, medial, and final timepoints. The left and right endpoints of the response scale for these items were labelled “Strongly Disagree” and “Strongly Agree,” respectively, with no labeled midpoint. The three sets of social closeness statements are shown in Table 2.2.

Domain	Set 1	Set 2	Set 3
Working	I would work with this person on a group project.	I would work on a team with this person.	I would work together with this person on an assignment.
Socializing	I would get coffee with this person.	I would have lunch with this person.	I would play board games with this person.
Living	I would be happy to have this person as a neighbor.	I would be happy to live next door to this person.	I would be happy if this person moved in to the house next to mine.

Table 2.2 Social closeness rating statements.

2.5 Language History and Debriefing Questionnaire

At the end of both experiments, participants completed a short questionnaire with questions about their native language, exposure to other languages, and their experience with the experiment. The full questionnaire is included in Appendix B. Participants who reported native languages other than English were excluded from data analysis. Participants who reported English and another language as native languages were included.

2.6 Participants

Data were collected from 48 undergraduate students at the University of Michigan. One participant reported “Chinese” as her sole native language in the language history questionnaire after the experiment, and her data were excluded from analysis. Three additional participants’ mouse-tracking data showed that they did not move the mouse on more than 40% of the 76 scale-response items and simply clicked through the question. This was interpreted as evidence of distraction or lack of engagement with the experiment, and their data were also excluded. Analyses were conducted on the data for the remaining 44 participants, most of whom failed to move the mouse on one or more scale-response items, but all of whom moved the mouse on at least 70% of items. Twenty-nine of these participants self-identified as women and 15 identified as men. All reported English as a native language; four reported an additional native language (these were Japanese, Russian, Arabic, and Urdu). All participants reported normal hearing and normal or corrected vision. Participants were paid $10 for completing the experiment.

2.7 Procedure and Data Collection

Participants were tested individually in a sound-attenuated booth in the Phonetics Lab at the University of Michigan. To allow accurate pupil diameter measurement, participants were seated with their chin and forehead resting on a desk-mounted support for the duration of the experiment, with a 90 second break after the medial block. The purpose of the study, as explained to participants, was to examine “how listeners perceive different speakers.” After providing informed consent, participants were given oral instructions to familiarize them with the procedure. Participants were told that the study involved several speakers with different ways of talking, and that the target speaker would be chosen for them randomly by the experiment software without the experimenter knowing which speaker they heard. This was intended to dissuade participants from trying to identify the speaker’s social group membership. Participants were also told that the speaker was reading the sentence stimuli, which were randomly ordered and unrelated to each other. They were instructed to type the words they heard, “even if the meaning seems odd or doesn't make sense," as quickly and accurately as was comfortable for them. For the intelligibility and attitude items, participants were told that they should answer each question according to their best judgment, even if they felt they did not have enough information. Participants were encouraged to be completely honest in their answers. All results were attached to anonymous participant numbers.

The experiment was carried out using SR Research Experiment Builder software. The language history and debriefing questionnaire was conducted in Qualtrics. Table 2.3 shows the stimuli and presentation order of the complete experiment in the test and control conditions. Participants in each condition were divided into three groups to partially counterbalance the order of the social perception and social closeness rating sets (see sections 2.4.2 and 2.4.3) to control for effects of set order that may have confounded the predicted change in attitude over time. The first group saw set 1 in the baseline and initial blocks, set 2 in the medial block, and set 3 in the final block. For the second group the order was: set 2, set 3, set 1; for the third group the order was: set 3, set 2, set 1. (Within each set, the order of the adjectives and closeness statements were the same for all participants in whichever block they were presented.) For all participants, the experiment was divided into four blocks: baseline, initial, medial, and final. Each block consisted of a sentence listening and transcription task followed by intelligibility, social perception (warmth and competence), and social closeness ratings, in that order. (In all four blocks, the same intelligibility items [see section 2.4.1] were presented in the same order for all participants.) In the baseline block, participants in both the test and control conditions heard the sentence stimuli spoken by the baseline speaker, who was a native speaker of American English. In the remaining blocks, participants in the test condition heard the target speaker in the non-native guise, while control participants heard the same speaker in the native guise.

Block	Task	Test			Control		
		CB1	CB2	CB3	CB1	CB2	CB3
Baseline	Transcription 10 sentences	Baseline speaker (American English accent)					
	Intelligibility rating	Intelligibility items					
	Social perception rating Social closeness rating	Item set 1	Item set 2	Item set 3	Item set 1	Item set 2	Item set 3
Initial	Transcription 10 sentences	Target speaker, Non-native guise			Target speaker, Native guise		
	Intelligibility rating	Intelligibility items					
	Social perception rating Social closeness rating	Item set 1	Item set 2	Item set 3	Item set 1	Item set 2	Item set 3
Medial	Transcription 40 sentences	Target speaker, Non-native guise			Target speaker, Native guise		
	Intelligibility rating	Intelligibility items					
	Social perception rating Social closeness rating	Item set 2	Item set 3	Item set 2	Item set 2	Item set 3	Item set 2
Break (90 seconds)							
Final	Transcription 40 sentences	Target speaker, Non-native guise			Target speaker, Native guise		
	Intelligibility rating	Intelligibility items					
	Social perception rating Social closeness rating	Item set 3	Item set 1	Item set 1	Item set 3	Item set 1	Item set 1
Language history and debriefing questionnaire							

Table 2.3 Experiment 1 presentation order and material. "CB" columns indicate partial counterbalancing of set presentation order. Each participant in both the test and control conditions was randomly assigned to one of three counterbalancing groups.

2.7.1 Objective Processing Fluency: Pupillometry and Sentence Transcription

Pupil diameter was captured with a remote monocular eye-tracker (EyeLink 1000 Plus, SR Research) sampling at 1000 Hz. After giving participants oral instructions, the experimenter performed a calibration procedure for tracking eye movement that, if necessary, was repeated until criterion was reached. The right pupil was measured except in a handful of cases when calibration failed for the right eye. The same pupil was measured throughout the experiment for each participant. To ensure consistent pupil measurement during the aural presentation of sentence stimuli, participants fixated on a cross in the center of the screen. Whenever a participant's gaze deviated more than 25 pixels from the cross, those measurements were omitted. Pupil diameter was measured from 1 second before the recorded sentence played until 1 second after it ended to capture accurate baseline and wrap-up measurements. Only the fixation cross was displayed on the screen during this period, and luminance was held constant for all trials to avoid confounding effects on pupil size. Sentence stimuli were heard over AKG 271 Mk2 headphones. After the pupil measurement period, participants typed what they heard in a text field on the screen. There was no time limit, but participants were instructed to transcribe as quickly and accurately as was comfortable for them.

2.7.2 Subjective Processing Fluency and Speaker Evaluation: Scale-Response Items

Participants recorded their ratings for the intelligibility, social perception, and social closeness items on a visually continuous scale using a marker controlled by the mouse. The scale was 640 pixels wide. The experiment software recorded the movement of the mouse as well as the horizontal pixel number where the participant clicked, which was used as the rating for that item. As described in section 2.4, the intelligibility and social closeness scales were labeled "Strongly Disagree" at the left endpoint and "Strongly Agree" at the right endpoint. The social perception scales were labeled "Not at all" at the left endpoint, "Somewhat" at the center (which was not otherwise marked), and "Very" at the right endpoint. There was no time limit for rating items.

2.8 Data Processing for Analysis

2.8.1 Pupil Dilation

The eye-tracking software used for measuring pupil diameter recorded measurements in arbitrary units that are not comparable between participants. Pupil diameter was first standardized for each participant, pooling data from the entire experiment to calculate mean and standard deviation. For each trial, the standardized pupil diameter over time curve was fit to a cubic smoothing spline with 40 degrees of freedom using the smooth.spline() function in the R *stats* package. Measurement samples that were missing due to blinks, saccades, or looks away from the fixation cross (annotated automatically by the Experiment Builder software) were interpolated using this spline. Figure 2.1 shows the raw pupil diameter curve and smoothed spline for a single sentence trial. This trial was chosen at random for the purposes of illustration and shows data for a participant in the Test condition listening to the sentence "My clock was wrong so I got to school late," in the non-native guise. Following recommendations based on previous research, trials with more than 20 percent of samples missing (10.9 percent of all trials) were excluded from analysis (Winn et al., 2018). For two participants in the control condition, this excluded more than 60% of total trials and more than 50% of trials in every block. Dilation results for these two participants were excluded from all dilation analyses. Pupil dilation in response to aural language stimuli has been shown to begin between 500 and 1300 ms after the onset of the stimulus. Accordingly, and consistent with other pupillometric research (Winn et al., 2018), the first 500 ms of diameter data for each sentence was used as baseline, and dilation was calculated using this mean baseline diameter subtracted from mean diameter in the 500 ms immediately after the sentence recording ended.

It should be noted that, while pupil dilation indexes cognitive effort, pupil constriction (i.e., a decrease in pupil diameter from the beginning to end of a stimulus) may not necessarily indicate cognitive ease. Dilation and constriction are largely controlled by separate physiological systems (Sirois & Brisson, 2014), and may respond differently to factors (such as luminance and cognitive effort) that influence pupil size. In tasks like sequence recall, constriction may indicate that a participant's cognitive capacity has reached its limit (Granholm et al., 1996). The literature consulted in designing and interpreting this experiment, including large-scale reviews of pupil dilation in speech processing (e.g., Winn et al., 2018; Zekveld et al., 2018), rarely mentions pupil constriction but appears to include constriction trials in analyses. Accordingly, the results presented in this chapter include both dilation and constriction trials. To investigate the potential ambiguity in the interpretation of pupil constriction, the main analyses of pupil dilation presented in sections 2.11.1 and 2.11.2 were also performed excluding constriction trials (i.e., data points below 0 on the y-axis in Figure 2.2 [top-left panel] and Figure 2.4 [left panel]), and the pattern of results was the same.

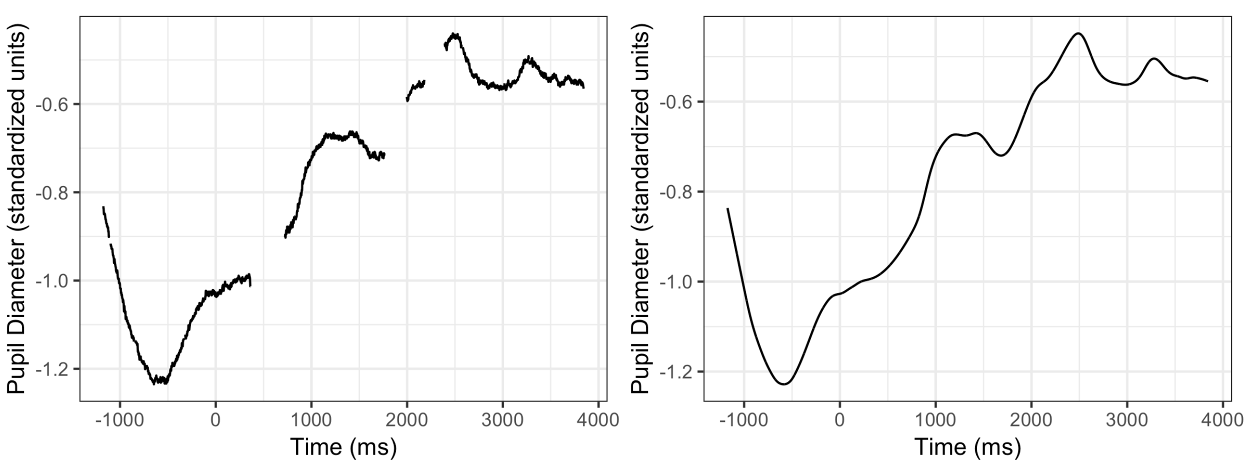

Figure 2.1 Raw (left) and smoothed (right) pupil diameter measurements, illustrating smoothing and interpolation for a single trial. Time is measured relative to the start of the sentence recording.

2.8.2 Transcription Errors

Sentence transcriptions were checked for typographical errors, misspellings, and other errors by the experimenter, who was blind to participant, condition, and trial order information. The following procedures were observed in identifying and correcting errors:

- Non-words that differed from real words by the addition or omission of one character, the transposition of two adjacent characters, or the substitution of one character with a keyboard-adjacent character were treated as typographical errors and corrected.
- Transcriptions with a homophonous word (e.g., “peace” for “piece”) were corrected.
- Commonwealth spellings (e.g., “favourite”) were corrected to United States spelling.
- Non-words that were most likely misspellings of a word that was not in the original sentence were corrected (e.g., “heros” to “heroes” and “possy” to “posse”).
- Compound non-words composed of two attested words were separated (e.g., “towelman” to “towel man”).
- In six of the trials, it was clear the participant had pressed "return" accidentally, intending to type an apostrophe, which resulted in an incomplete transcription. These trials were excluded from analysis.

All other differences from the original written sentence were treated as intentional and left as transcribed, even if the difference resulted in a nonsensical sentence. When necessary, pronunciations were created for non-words by the experimenter based on U.S. English spelling conventions. These words and pronunciations are listed in Appendix C.

The original sentences and transcriptions were converted to the International Phonetic Alphabet using the Carnegie Mellon Pronouncing Dictionary. Word breaks were represented by the # symbol and counted as a phone for scoring. Each transcription was compared to the original sentence to compute the phone edit distance between the two strings of IPA symbols. This measure counts the number of symbol insertions, deletions, or substitutions required to make the transcription match the original sentence.

2.8.3 Scale-Response Item Scoring

Scale rating responses ranged from 1 to 641, corresponding to the number of pixels in the scale on the screen. The ratings for the social perception adjectives with negative favorability reported by Bergsieker et al. (2012; see table 2.1) were reverse scored for analysis so that, for all social perception item ratings, 1 corresponds to the most negative evaluation of the speaker, and 641 corresponds to the most positive. The intelligibility and social closeness items were all designed with positive valence and raw scores were used in the analyses.

The intelligibility, warmth, competence, and closeness ratings in the analyses represent the average of ratings for the items in each dimension, unless otherwise noted. For intelligibility, internal consistency across the three items was high in all four blocks (Baseline block $\alpha = 0.88$, Initial block $\alpha = 0.96$, Medial block $\alpha = 0.96$, Final block $\alpha = 0.98$). Calculating Cronbach's α is not possible for the other dimensions because different groups of participants rated different items in each block. To account for variation among the items in each dimension, a random intercept by Item was included in the respective models. (Full descriptions of the models are provided in sections 2.10.1 and 2.10.2.)

2.8.4 The Relationship Between Fluency and Attitude and Change over Time

In order to compare intelligibility and attitude ratings (which were measured once per block) with dilation, edit distance, and transcription time (which were measured for each transcription trial), summary statistics for the latter variables were calculated using the following procedure. For the final 10 transcription trials in each block (i.e., trials immediately preceding the scale-response ratings), Dilation, Edit Distance, and Transcription Time for each sentence were divided by the number of phones in the sentence, in order to control for the effects of sentence length. The adjusted value was multiplied by 31, the average sentence length in phones, to aid the interpretation of model estimates. The mean of the 10 length-adjusted values provides the summary statistic for comparison with Intelligibility, Warmth, Competence, and Closeness rating. To compare changes in the fluency and rating variables over time, the difference of means between the initial and medial blocks and between the medial and final blocks were used as the change summary statistic.

2.9 Predictions

2.9.1 Processing Fluency

The first main hypothesis tested here is that objective processing fluency (measured as pupil dilation, transcription errors, transcription time) and subjective processing fluency (ratings on intelligibility items) will increase over time, and will increase to a greater degree for the participants in the test (compared to the control) condition, as they gain experience with the non-native accent. This hypothesis predicts that Dilation, Edit Distance, and Transcription Time will decrease over trials, and Intelligibility ratings will increase from the Initial to Medial block and from the Medial to Final block. These changes are predicted to be larger in the Test condition than the Control condition. These predictions are also consistent with the secondary hypothesis that increased accuracy over time in transcribing a non-native speaker is due to increased processing fluency, rather than increased effort on the part of the participant. While previous research has not found a reliable correlation between pupil dilation and intelligibility rating (see Winn et al., 2015), the relationships among dilation, accuracy, transcription time, and intelligibility will be tested. If participants are subjectively aware of increasing fluency over time, the means of Dilation, Edit Distance, and Transcription Time over the 10 transcription trials immediately preceding intelligibility rating are predicted to negatively correlate with ratings for the three Intelligibility items.

Based on previous research, processing fluency is also expected to be lower for a non-native accent than a native accent. Accordingly, Dilation, Edit Distance, and Transcription Time are predicted to increase from the Baseline to Initial block more for participants in the Test condition than the Control condition. Similarly, Intelligibility ratings are predicted to decrease from the Baseline to Initial block more in the Test condition than the Control condition. For the Initial, Medial, and Final blocks, Dilation, Edit Distance, and Transcription Time are predicted to be higher on average in the Test condition than the Control condition; Intelligibility ratings are predicted to be lower in the Test condition than the Control condition.

2.9.2 Speaker Evaluation

The second main hypothesis is that attitudes toward the speaker will improve with experience with a non-native accent. This hypothesis predicts that ratings for the Warmth, Competence, and Closeness items will increase from the Initial to Medial block and from the Medial to Final block to a greater degree in the Test condition than the Control condition. The hypothesized relationship between attitude and processing fluency predicts that mean Dilation, Edit Distance, and Transcription Time for the 10 transcription trials immediately preceding each rating will negatively correlate with Warmth, Competence, and Closeness rating. Intelligibility rating is predicted to positively correlate with attitude rating.

Based on previous research, social evaluation is also expected to be more negative for a non-native accent than a native accent. Accordingly, Warmth, Competence, and Closeness ratings are predicted to decrease from the Baseline to Initial block more in the Test condition than the Control condition. For the Initial, Medial, and Final blocks, ratings are predicted to be lower in the test condition than in the control condition.

2.9.3 The Relationship Between Fluency and Evaluation

Improvement in attitudes over time is hypothesized to depend on the improvement in processing fluency. Accordingly, the differences of means of Dilation, Edit Distance, and Transcription Time between the Initial and Medial blocks and between the Medial and Final blocks are predicted to negatively correlate with differences in Warmth, Competence, and Closeness rating. The Initial-Medial and Medial-Final differences in Intelligibility rating are predicted to positively correlate with the corresponding differences in Warmth, Competence, and Closeness rating.

2.10 Statistical Analysis

Predictions for all variables except Edit Distance were tested using linear models, including random effects as necessary using the R *lme4* package. In models with random effects, *p*-values were computed using Satterthwaite's degrees-of-freedom method using the R *lmerTest* package. Predictions for Edit Distance were tested using generalized linear mixed-effects models (LMEMs) with a Poisson-distributed dependent variable and a log link function, also using the R *lme4* package. Standard errors for estimates in the Edit Distance models were adjusted by the overdispersion ratio to calculate quasi-likelihood estimates to avoid anticonservative *p*-values. When necessary, post-hoc tests of targeted contrasts were performed using the R *multcomp* package, with *p*-values adjusted using the single-step method. Because of the large number of comparisons used to test the predictions described in section 2.9, significance should be interpreted with care. However, most *p*-values reported in section 2.11 are less than 0.01, indicating that the overall pattern of results is likely reliable. Comparisons that yield larger *p*-values are noted in the presentation and discussion of the results.

The distribution of Transcription Time was positively skewed and was $\log_{10}$-transformed when tested as a dependent variable to avoid violating model assumptions. (Transcription Time was measured in milliseconds and a base-10 logarithm was used to aid interpretation of model results and data visualizations.) In all regression models that included the predictor variable Condition (Test vs. Control), the contrast values used simple coding[3] with Control as the reference level. The β-estimates for the main effect of Condition in these models, therefore, indicate the difference between the Test and Control condition means.

---

[3] Simple coding (also referred to as "sum coding") for a two-level categorical variable like Condition assigns a numerical value of -0.5 to the reference level (Control) and 0.5 to the comparison level (Test). The *β*-estimate for a main effect of Condition indicates a model-predicted mean change from the reference level to the comparison level, i.e., the difference of Test - Control. Because the sum of values for the two levels is 0, the main effect *β*-estimate for a second model predictor variable *y* can be interpreted without regard to the value of Condition (i.e., Test vs. Control), unless there is an interaction between *y* and Condition. *β*-estimates for interactions between Condition and *y* are interpreted as the change in *y* when Condition = 0, i.e., the model-predicted mean change from the reference level of *y* to the comparison level of *y* on average for the Control and Test conditions together. If *y* is also a two-level, simple-coded categorical variable, the interaction can be interpreted as the effect of Condition when *y* = 0, i.e., the change from Control to Test condition on average for both values (i.e., levels) of *y* together.

2.10.1 Processing Fluency and Attitudes for Native vs. Non-Native Accents

Differences in Dilation, Edit Distance, Transcription Time, and the scale-response ratings between the Baseline and Initial blocks were tested using models with Block (Baseline vs. Initial), Condition (Test vs. Control), Sentence Length (in phones), and the interaction of Block and Condition as fixed effects, with a random intercept by Participant. Models with scale-response ratings as the dependent variable also included a random intercept by rating Item. The Block variable contrast used simple coding with Baseline as the reference level. The main effect of Condition in the models described in section 2.10.2 also tested differences in processing fluency and attitude depending on accent.

2.10.2 Change in Fluency and Attitude with Experience

For Dilation, Edit Distance, and Transcription Time in the Initial, Medial, and Final blocks, each variable was tested in a model with Trial Order (an integer ranging from 1 to 90), Condition (Test vs. Control), Sentence Length (in phones) and the interaction of Trial Order and Condition as fixed effects. Differences in scale-response ratings between the Initial, Medial, and Final blocks were tested using models with Block (Initial vs. Medial, and Medial vs. Final), Condition (Test vs. Control), and their interaction as fixed effects. The Block variable contrast used backward-difference coding to compare the Medial block to the Initial blocks and the Final block to the Medial block. In this coding system, the Final and Initial blocks are not directly compared.[4] All models included a random intercept by Participant. Models with Intelligibility, Warmth, Competence, or Closeness rating as the dependent variable also included a random intercept by Item.

2.10.3 The Relationship Between Fluency and Attitudes

The effect of processing fluency on Warmth, Competence, and Closeness was tested in LMEMs using data from the Initial, Medial, and Final blocks together. Separate models were used for each attitude dimension. Each model was constructed by adding predictor variables (mean Dilation, mean Edit Distance, mean Transcription Time, and Intelligibility), one at a time, as fixed effects to a base model with a fixed effect of Condition and a random intercept by Participant. ANOVA model comparisons were used to test whether the addition of each effect significantly improved the model fit, and predictor variables that did not were omitted from the final model. Interactions of the predictor variables with Condition (Test vs. Control) were also added and tested.

2.10.4 Change in Fluency and Attitudes

The relationship between changes in processing fluency and changes in evaluation were tested in linear models with one data point per participant and did not include random effects.

---

[4] In backward-difference coding of a three-level categorical variable such as Block, the Initial, Medial, and Final levels of the variable are assigned different numerical values for the Initial-Medial contrast and the Medial-Final contrast. Because the sums of numerical values assigned to the levels for both contrasts equal 0, the main effect β-estimate of a second variable such as Condition can be interpreted without regard to the value (i.e., Initial, Medial, or Final) of Block, unless Block and Condition interact. β-estimates for the interaction between Condition and the Initial-Medial contrast are interpreted as the change in the difference between the Test and Control conditions (i.e., Test - Condition) from Initial to Medial block. The interaction between Condition and the Initial-Medial contrast can also be interpreted as the Initial-to-Medial change in the Test group minus Initial-to-Medial change in the Control group. The interaction between Condition and the Medial-Final contrast is similarly interpreted.

Six models were constructed altogether for Initial-Medial and Medial-Final rating differences for Warmth, Competence, and Closeness. Each model was constructed by the same iterative fixed-effect addition procedure described in section 2.10.3 using differences in mean Dilation, mean Edit Distance, mean Transcription Time, and Intelligibility as predictor variables.

2.10.5 The Relationship Between Fluency Measures

The relationship of pupil dilation as a physiological measure of processing fluency with transcription accuracy and transcription time as behavioral measures of processing fluency was tested using two separate models. Edit Distance was regressed over Dilation in a Poisson GLMEM and Transcription Time was regressed over Dilation in an LMEM, both with random intercepts by Participant. The model of Edit Distance included fixed effects of Condition and its interaction with Dilation. For the Transcription Time model, analysis of residuals indicated that the relationship of the dependent variable with Dilation was non-linear, and an orthogonal second-order polynomial effect of Dilation significantly improved model fit. Condition was included as a fixed effect but its interactions with the first- and second-order effects of Dilation did not significantly improve model fit and were omitted from the final model. The effects of Dilation, Edit Distance, and Transcription Time (objective measures of fluency) on Intelligibility (a subjective measure of fluency) were also tested. The design of models testing these effects was influenced by the results described in sections 2.11.3 and 2.11.4, and the model construction and selection procedure will be described in section 2.11.5.

2.11 Results

2.11.1 Processing Fluency and Attitudes for Native vs. Non-Native Accents

Processing was predicted to be more fluent and speaker evaluation more positive for native-accented speech (i.e., in the Baseline block for all participants and in the Initial, Medial, and Final blocks for Control participants) than non-native accented speech (i.e., in the Initial Medial and Final blocks for Test participants). Comparison of results in the Baseline and Initial blocks upheld this prediction for the dependent variables Dilation, Edit Distance, and Transcription Time (objective measures of fluency); Intelligibility (subjective measure of fluency); and Competence (one of three attitude dimensions). Results for Closeness ratings were less clear, and the prediction was not upheld for Warmth. Data for these comparisons are shown in Figure 2.2 and Figure 2.3. Model parameters and statistics for the Baseline-Initial comparison are shown in Table 2.4.

For Dilation and Edit Distance, the significant main effect of Initial Block and the significant interaction between Condition and Block indicate that each variable increased from the Baseline to Initial block more in the Test condition than in the Control condition, reflecting less fluent processing of the non-native accent than the native accent (Figure 2.2, top-left and bottom-left panels). The patterns are the same for Transcription Time (Figure 2.2, top-right panels) but the effects are only marginally significant. For Intelligibility (Figure 2.2, bottom-right panels) and Competence (Figure 2.3, middle panels), the significant effects of Block and the interaction of Condition and Block indicate that ratings decreased from the Baseline to Initial block more in the Test condition than the Control condition.

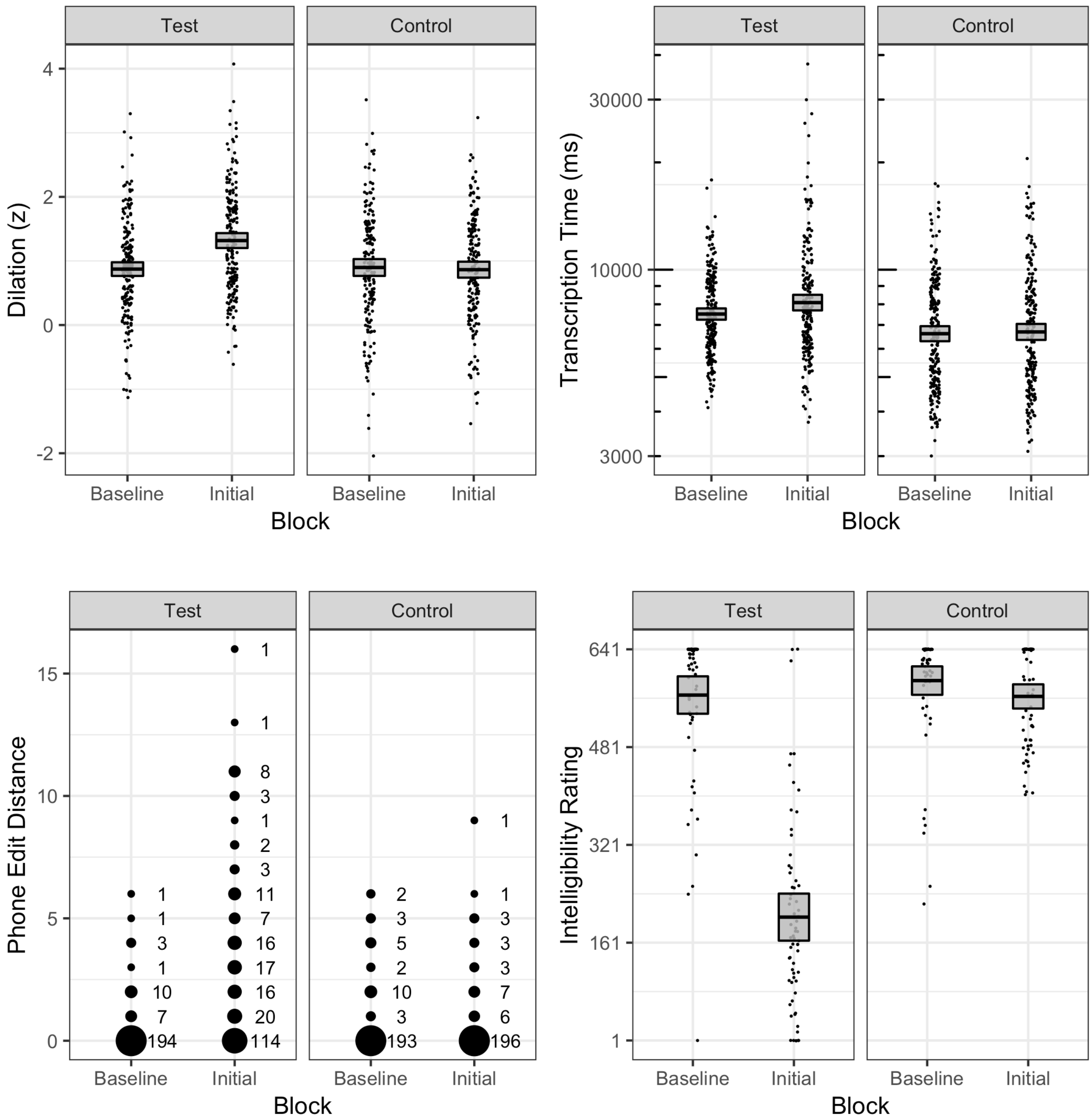

Figure 2.2 Dilation (top left), Transcription Time (top right), Edit Distance (bottom left), and Intelligibility ratings (bottom right; see section 2.4.1 for rating items) in the Baseline and Initial blocks. Transcription Time is displayed on a $\log_{10}$-adjusted scale. Cross bars indicate the mean and 95% confidence intervals. In the plot of Edit Distance, circle area is proportional to the number of observations of each value, with the number annotated next to each point. The effect of Block and the interaction of Block and Condition were significant for Dilation, Edit Distance, and Intelligibility ratings and marginally significant for Transcription Time.

For Closeness (Figure 2.3, right panels), the main effect of Block (β = 64.620, t = 3.935, p < 0.001) and the interaction of Condition and Block (β = -115.284, t = -3.509, p = 0.001) indicate that ratings increased from Baseline to Initial block more in the Control condition than the Test condition. While this pattern is not contrary to predictions, the increase in ratings in the Control condition is unexpected. If Closeness ratings depend in part on voice quality, this result may indicate that listeners preferred the target speaker to the baseline speaker, and in the Test condition this preference was counteracted by the non-native accent. For Warmth (Figure 2.3, left panels), the main effect of Block (β = 62.610, t = 6.201, p < 0.001), with no significant interaction of Condition and Block, indicates that ratings increased from the Baseline to Initial blocks with no difference between conditions.

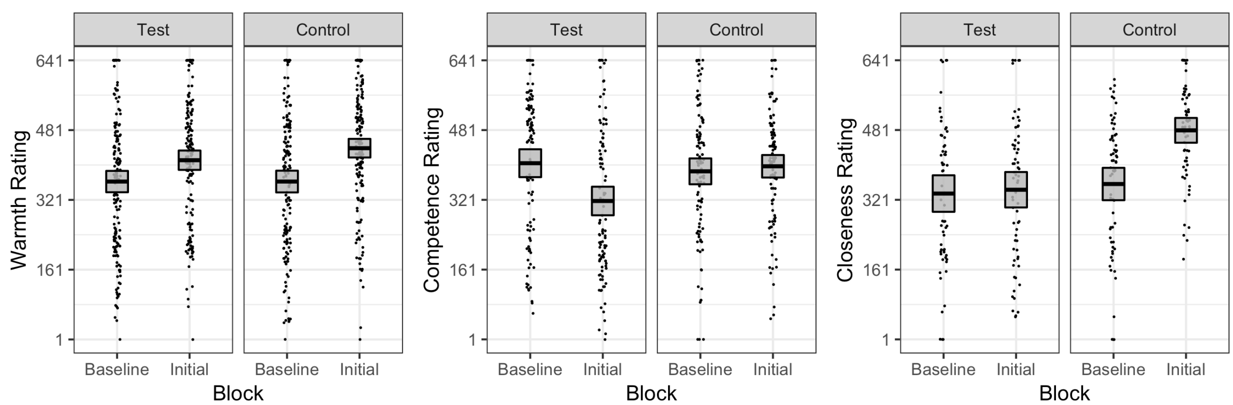

Figure 2.3 Warmth (left), Competence (center), and Closeness (right) ratings in the Baseline and Initial blocks. (See Table 2.1 and Table 2.2 for lists of items.) Cross bars indicate the mean and 95% confidence intervals. The effect of Block and the interaction of Condition and Block were significant for Competence and Closeness. The effect of Block was significant for Warmth, but there was no significant interaction of Condition and Block.

Dependent Variable	Fixed Effect (Predictor Variable)	β	Test Statistic[a]	p
Dilation (z)	**(Intercept)**	**0.684**	**5.550**	**< 0.001**
	Initial Block	**0.187**	**3.779**	**< 0.001**
	Test Condition	0.221	1.629	0.111
	Sentence Length (phones)	**0.009**	**2.668**	**0.008**
	Block * Condition	**0.498**	**5.019**	**< 0.001**
Transcription Time ($\log_{10}$ ms)	**(Intercept)**	**3.500**	**153.934**	**< 0.001**
	Initial Block	**0.018**	**2.565**	**0.011**
	Test Condition	*0.067*	*1.915*	*0.062*
	Sentence Length (phones)	**0.011**	**24.040**	**< 0.001**
	*Block * Condition*	*0.027*	*1.945*	*0.052*
Edit Distance	**(Intercept)**	**-2.508**	**-7.844**	**< 0.001**
	Initial Block	**0.933**	**5.516**	**< 0.001**
	Test Condition	**0.982**	**2.585**	**0.010**
	Sentence Length (phones)	**0.052**	**7.048**	**< 0.001**
	Block * Condition	**2.006**	**5.930**	**< 0.001**
Intelligibility Rating	**(Intercept)**	**481.101**	**37.538**	**< 0.001**
	Initial Block	**-193.738**	**-17.534**	**< 0.001**
	Test Condition	**-192.220**	**-7.499**	**< 0.001**
	Block * Condition	**-337.167**	**-15.258**	**< 0.001**
Warmth Rating	**(Intercept)**	**392.061**	**27.405**	**< 0.001**
	Initial Block	**62.610**	**6.201**	**< 0.001**
	Test Condition	-14.951	-0.738	0.465
	Block * Condition	-30.092	-1.491	0.136
Competence Rating	**(Intercept)**	**375.895**	**20.238**	**< 0.001**
	Initial Block	**-37.889**	**-2.937**	**0.004**
	Test Condition	-28.800	-1.307	0.198
	Block * Condition	**-99.076**	**-3.840**	**< 0.001**
Closeness Rating	**(Intercept)**	**379.080**	**25.576**	**< 0.001**
	Initial Block	**64.620**	**3.935**	**< 0.001**
	Test Condition	**-78.540**	**-2.899**	**0.006**
	Block * Condition	**-115.284**	**-3.509**	**0.001**

Table 2.4 Model parameters and statistics comparing Baseline and Initial blocks. Bold effects are significant at $p < .05$. Italic effects are marginally significant at $p < .1$.

[a] For the Poisson GLMEM of Edit Distance, the test statistic was *z*. For all other models, the test statistic was *t*.

2.11.2 Change in Fluency and Attitude with Experience

Both subjective and objective processing fluency were predicted to improve over time, with a larger improvement in the Test condition (compared to the Control condition) as participants gained experience with the non-native accent. Results for the objective fluency measures of Dilation and Transcription Time (Figure 2.4) are consistent with this prediction. Dilation (Figure 2.4, left panels) decreased significantly across trials ($\beta = -0.006$, $t = -13.149$, $p < 0.001$), and there was a significant interaction between Trial Order and Condition ($\beta = -0.003$, $t = -3.776$, $p < 0.001$), indicating that the decrease was larger in the Test condition than in the Control condition.[5] The pattern of effects was the same for Transcription Time (Figure 2.4, right panels) with a significant decrease across trials ($\beta = -2.2 \times 10^{-4}$, $t = -3.512$, $p < 0.001$), and a significant interaction between Trial Order and Condition ($\beta = -4.1 \times 10^{-4}$, $t = -3.207$, $p = 0.001$).[6] Results for the objective fluency measure Edit Distance (Figure 2.5), though, were not consistent with predictions. The absence of a significant effect of Trial Order or its interaction with Condition may be due to near-ceiling accuracy in both conditions; Edit Distance was 0 (indicating no difference between the recorded sentence and the transcription) in 63.4% of Test trials and 89.7% of Control trials. The significant positive main effects of Condition on Dilation ($\beta = 0.387$, $t = 3.353$, $p = 0.001$), Edit Distance ($\beta = 2.013$, $z = 5.558$, $p < 0.001$) and Transcription Time ($\beta = 0.081$, $t = 2.232$, $p = 0.031$) are consistent with the

[5] The β estimates in this model indicate the effect of trial order on Dilation measured in standardized units. Assuming a normally distributed pupil diameter range of 3 mm (based on data reported by Richman et al., 2004), the model indicates that average dilation is about .19 mm larger in the test condition than in the control. With the same assumptions, the decrease over trials is about .35 mm in the test condition and about .19 mm in the control condition

[6] In real terms, the effect of Trial Order was extremely small, with average Transcription Time in the Test condition decreasing by only about 3 ms over the experiment.

prediction that processing is more fluent for native than non-native accented speech. A full summary of parameter estimates and statistics for the relevant LMEMs is given in Table 2.5.

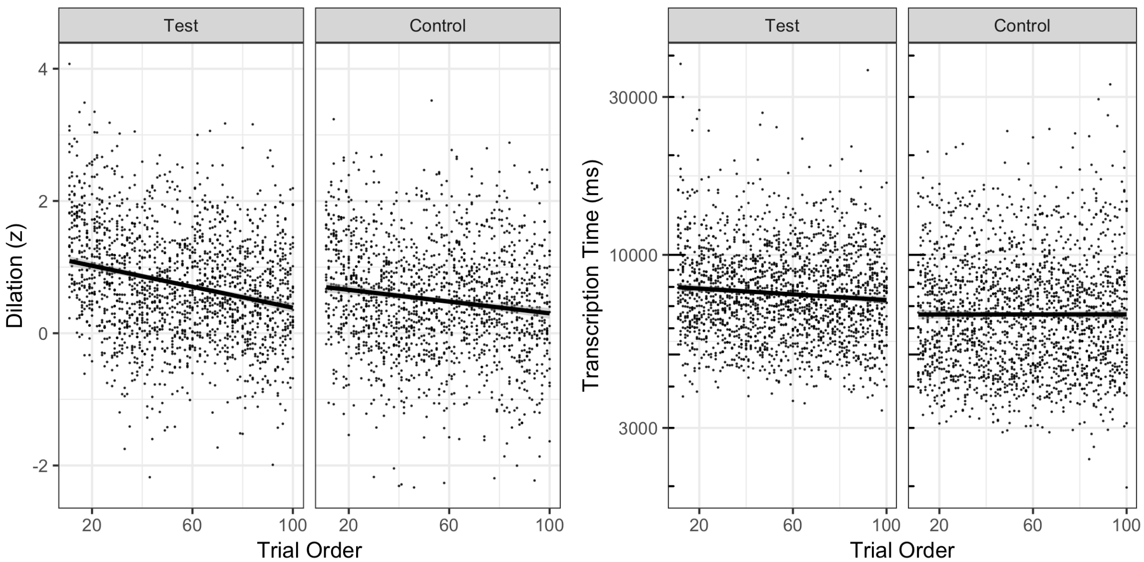

Figure 2.4 Dilation (left) and Transcription Time (right) by Trial Order in the Initial (trials 11-20), Medial (trials 21-60), and Final (trials 61-100) blocks. Transcription Time is plotted on a $\log_{10}$-transformed scale. The effect of Trial Order and the interaction of Trial Order and Condition were significant for both dependent variables.

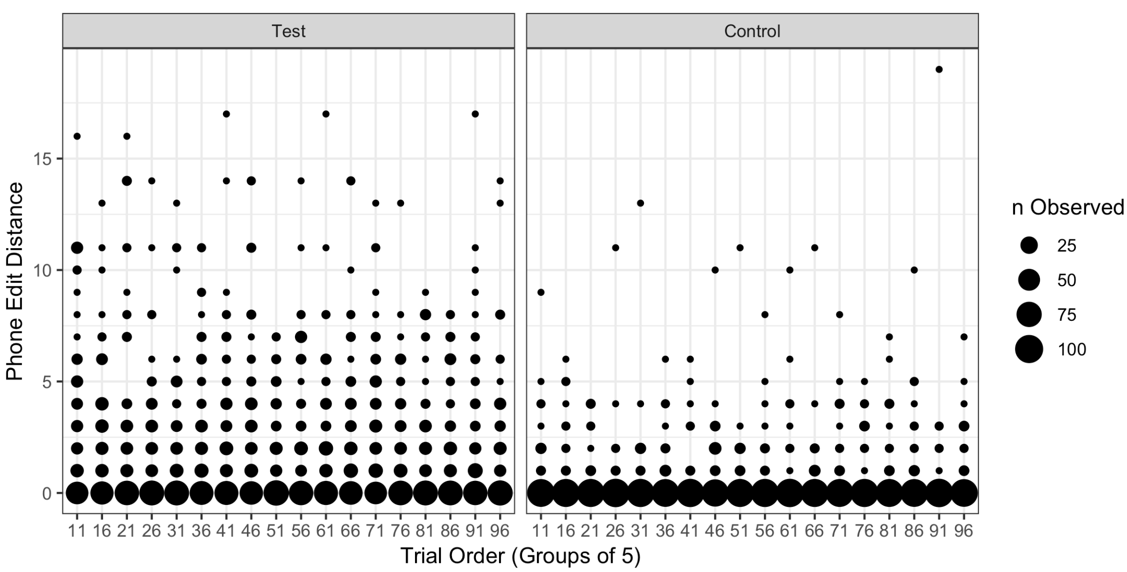

Figure 2.5 Edit Distance by Trial Order in the Initial (trials 11-20), Medial (trials 21-60), and Final (trials 61-100) blocks. Edit Distance was significantly higher in the Test condition than the Control condition. The effects of Trial Order and the interaction of Trial Order and Condition were not significant.

Results for Intelligibility ratings (Figure 2.6), a measure of subjective fluency, were consistent with the prediction that participants would perceive processing to be easier for a native accent than a non-native accent, as indicated by the significant main effect of Condition ($\beta = -349.360$, $t = -12.854$, $p < 0.001$). However, the predictions that Intelligibility ratings would increase from Initial to Medial and from Medial to Final blocks, with a greater increase in the Test condition than the Control condition, were not upheld. In the model comparing ratings by Condition and Block, there was not a significant change from Initial to Medial block or interaction with Condition. Mean ratings did, though, decrease significantly from the Medial to the Final block ($\beta = -24.705$, $t = -2.383$, $p = 0.018$), and this effect showed a significant interaction with Condition ($\beta = 53.016$, $t = 2.557$, $p = 0.011$). Post-hoc tests indicated that ratings decreased significantly in the Control condition ($\beta = -51.213$, $t = -3.476$, $p = 0.001$) but did not change significantly in the Test condition. Parameter estimates and statistics for the model are shown in Table 2.5. Possible interpretations of the unexpected decrease in Intelligibility ratings in the Control condition but not the Test condition will be discussed in sections 2.12 and 4.2.1.

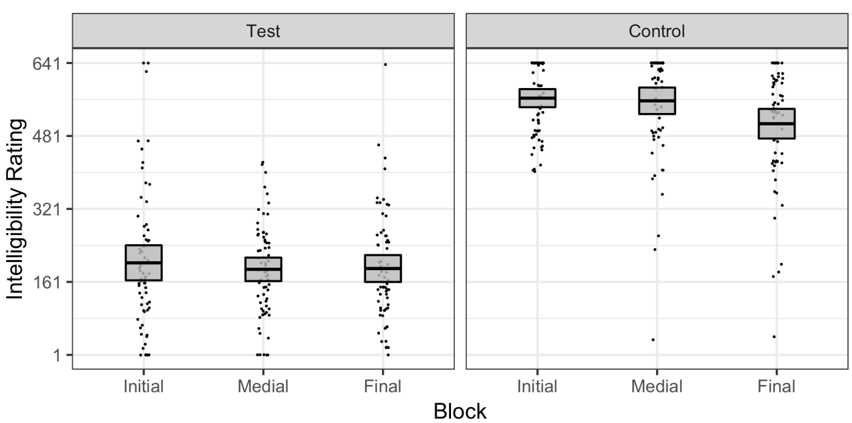

Figure 2.6 Intelligibility ratings by Block and Condition. Ratings decreased significantly from the Medial to Final blockn the Control but not the Test condition. No significant differences were found between Initial and Medial blocks.

Dependent Variable	Fixed Effect (Predictor Variable)	β	Test Statistic[a]	p
Dilation (z)	**(Intercept)**	**0.797**	**10.461**	**< 0.001**
	Trial Order	**-0.006**	**-13.149**	**< 0.001**
	Test Condition	**0.387**	**3.353**	**0.001**
	Sentence Length (phones)	0.002	1.546	0.122
	Trial Order * Condition	**-0.003**	**-3.776**	**< 0.001**
Transcription Time (log_{10} ms)	**(Intercept)**	**3.503**	**179.844**	**< 0.001**
	Trial Order	**2.2×10^{-4}**	**-3.512**	**< 0.001**
	Test Condition	**0.081**	**2.232**	**0.031**
	Sentence Length (phones)	**0.011**	**51.667**	**< 0.001**
	Trial Order * Condition	**4.1×10^{-4}**	**-3.207**	**0.001**
Edit Distance	**(Intercept)**	**-1.795**	**-7.963**	**< 0.001**
	Trial Order	-0.001	-0.544	0.586
	Test Condition	**2.013**	**5.558**	**< 0.001**
	Sentence Length (phones)	**0.042**	**10.586**	**< 0.001**
	Trial Order * Condition	-0.005	-1.535	0.125
Intelligibility Rating	**(Intercept)**	**369.210**	**20.873**	**< 0.001**
	Block 1 Initial to Medial	-10.492	-1.006	0.315
	Block 2 Medial to Final	**-25.062**	**-2.403**	**0.017**
	Test Condition	**-348.475**	**-12.826**	**< 0.001**
	Block 1 * Condition	-6.288	-0.302	0.763
	Block 2 * Condition	**52.340**	**2.509**	**0.013**

Table 2.5 Model parameters and statistics for comparisons across trials/blocks. Bold effects are significant at $p < .05$.

[a] For Edit Distance, the test statistic was *z*. For all other models, the test statistic was *t*.

Warmth, Competence, and Closeness ratings in the Initial, Medial, and Final blocks are shown in Figure 2.7. Parameter estimates and statistics for the corresponding models are summarized in Table 2.6. Ratings on all three dimensions were predicted to increase from Initial to Medial block and from Medial to Final block, with larger increases in the Test condition than the Control condition. The results for Warmth and Closeness were inconsistent with this prediction, while the prediction for Competence was partly upheld.

Warmth ratings (Figure 2.7, top-left panels) decreased significantly from the Initial to Medial blocks (β = -23.878, t = -2.310, $p = 0.021$), but no other effects of Condition, Block, or their interaction were significant. The absence of a main effect of Condition on Warmth was contrary to predictions but consistent with the absence of a significant main effect of Condition in the model comparing Warmth ratings in the Baseline and Initial blocks (see section 2.11.1; Figure 2.3, left panels; and Table 2.4).

In the model of Competence ratings (Figure 2.7, top-right panels), there were no significant main effects of Block, but there was a significant main effect of Condition (β = -50.227, t = -2.233, $p = 0.031$), consistent with the prediction that the speaker would be evaluated as more competent in the native (Control) guise than the non-native (Test) guise. Condition significantly interacted with Initial-Medial Block (β = 55.536, t = 2.258, $p = 0.024$), and post-hoc tests indicated that there was a marginally significant decrease in Competence ratings from the initial to medial block in the Control condition (β = -34.759, t = -1.913, $p = 0.063$) but no significant change in the Test condition. The marginal significance of the decrease in the Control condition may indicate that this pattern of results was observed by chance. This effect will be tested again with a different group of participants in Experiment 2.

In the model of Closeness ratings (Figure 2.7, bottom panels), there was a significant main effect of Condition (β = -106.887, t = -3.255, $p = 0.002$), consistent with lower ratings for the non-native guise than the native guise, as predicted. There were no significant main effects of Block, but there was a marginally significant interaction between Condition and Medial-Final Block (β = 47.475, t = 1.890, $p = 0.060$), indicating that the change from Medial to Final block was more positive in the Test condition than the Control condition. Post-hoc tests indicated that the Medial-Final change was not significantly different from zero in either condition.

This significant main effect of Condition on Closeness ratings (with only a marginally significant interaction with Block) supports the prediction that ratings would be lower for the non-native accent than the native accent and may clarify the analysis of Closeness ratings in the Baseline and Initial blocks (see 2.11.1 and Table 2.4). That test showed that ratings increased from Baseline to Initial block more for the Control group than the Test group. The two results together suggest that some factor other than accent (such as voice quality) influenced different Closeness ratings for the baseline speaker and the target speaker, and that target speaker evaluation was negatively influenced by accent for the non-native guise but not the native guise.

Dependent Variable	Fixed Effect (Predictor Variable)	β	t	p
Warmth	**(Intercept)**	**410.803**	**28.952**	**< 0.001**
	Block 1 Initial to Medial	**-23.878**	**-2.310**	**0.021**
	Block 2 Medial to Final	6.352	0.513	0.608
	Test Condition	-20.036	-0.976	0.335
	Block 1 * Condition	10.793	0.557	0.577
	Block 2 * Condition	3.522	0.182	0.856
Competence	**(Intercept)**	**355.433**	**20.517**	**< 0.001**
	Block 1 Initial to Medial	-6.991	-0.534	0.593
	Block 2 Medial to Final	9.535	0.610	0.542
	Test Condition	**-50.227**	**-2.233**	**0.031**
	Block 1 * Condition	**55.536**	**2.258**	**0.024**
	Block 2 * Condition	-26.603	-1.087	0.277
Closeness	**(Intercept)**	**400.492**	**24.207**	**< 0.001**
	Block 1 Initial to Medial	-14.900	-1.177	0.240
	Block 2 Medial to Final	-4.863	-0.375	0.709
	Test Condition	**-106.887**	**-3.255**	**0.002**
	Block 1 * Condition	17.739	0.705	0.481
	*Block 2 * Condition*	*47.475*	*1.890*	*0.060*

Table 2.6 Model statistics and parameters for Warmth, Competence, and Closeness comparing Initial, Medial, and Final blocks. Bold effects are significant at $p < 0.05$. Italic effects are marginally significant at $p < 0.1$.

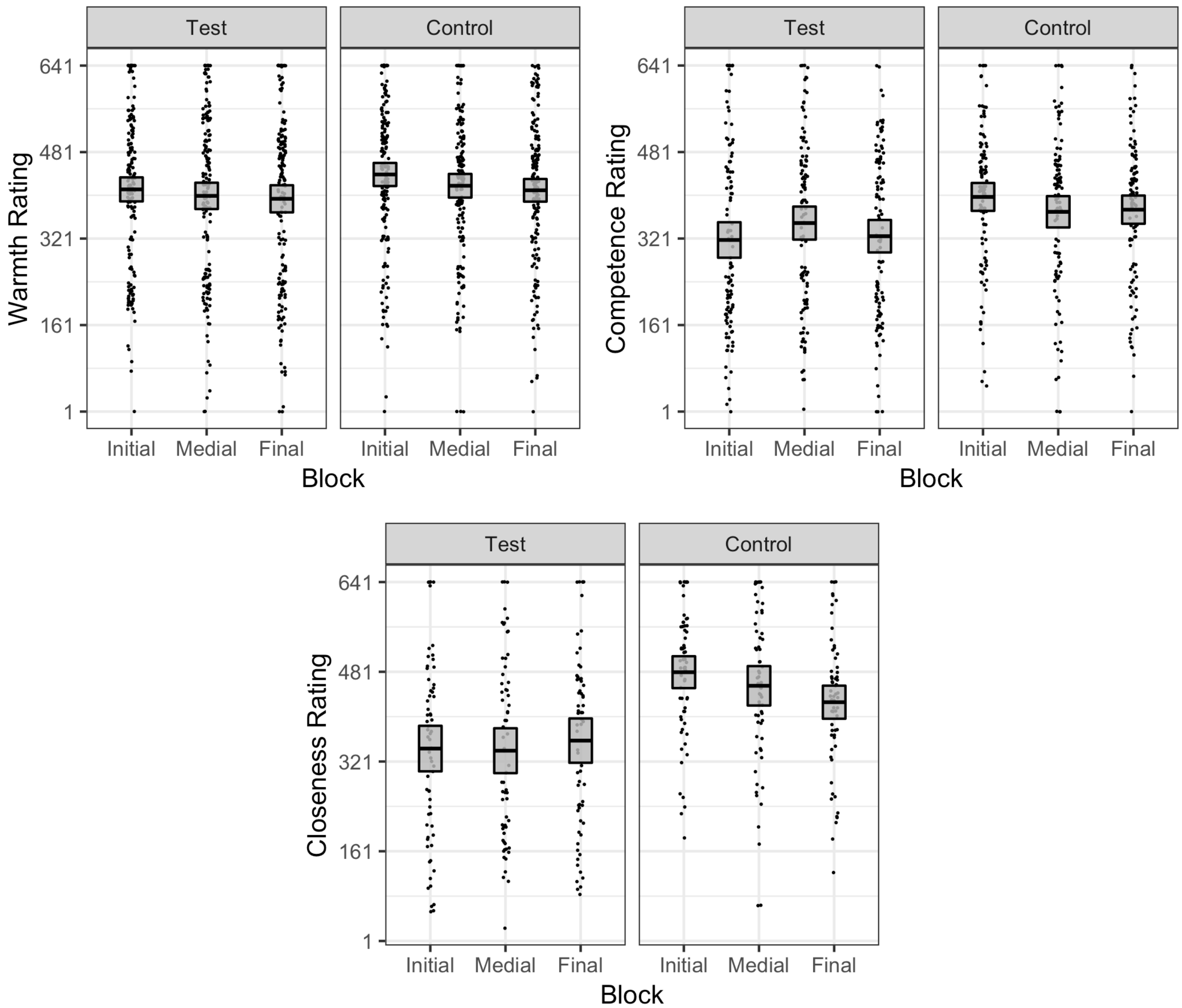

Figure 2.7 Warmth (top left), Competence (top right), and Closeness (bottom) ratings by Block and Condition. Warmth ratings decreased significantly from the Initial to Medial blocks. Competence and Closeness ratings were significantly lower in the Test condition than the Control condition. For Competence ratings, there was a significant interaction of Initial-Medial Block and Condition. For Closeness ratings, there was a marginally significant interaction of Medial-Final Block and Condition.

2.11.3 The Relationship Between Fluency and Attitudes

Objective processing fluency (measured using Dilation, Edit Distance, Transcription Time) and subjective processing fluency (measured using Intelligibility rating) were predicted to correlate positively with attitude rating on the Warmth, Competence, and Closeness dimensions.

The results of the model construction and selection procedure testing the effects of fluency on attitudes (summing across data from the Initial, Medial, and Final blocks) were largely inconsistent with this prediction. There were no significant correlations of Dilation, Edit Distance, or Transcription Time with attitude rating. Intelligibility rating (Figure 2.8), the measure of subjective fluency, had a significant positive correlation with Warmth ($\beta = 0.132$, $t = 1.979$, $p = 0.049$), Competence ($\beta = 0.191$, $t = 2.404$, $p = 0.018$), and Closeness ratings ($\beta = 0.378$, $t = 4.098$, $p < 0.001$). However, the main effect of Condition and its interaction with Intelligibility on attitude rating were not significant. Parameter estimates and statistics for models testing these relationships are shown in Table 2.7. These results support the prediction that the subjective experience of greater ease is associated with more positive speaker evaluation, but suggest that the subjective experience of ease may not reflect objective fluency. An analysis of the relationship between objective and subjective fluency is presented in section 2.11.5.

Dependent Variable	Fixed Effect (Predictor Variable)	β	t	p
Warmth Rating	**(Intercept)**	**366.554**	**12.706**	**< 0.001**
	Intelligibility Rating	**0.132**	**1.979**	**0.049**
	Test Condition	21.702	0.651	0.517
Competence Rating	**(Intercept)**	**285.570**	**9.153**	**< 0.001**
	Intelligibility Rating	**0.191**	**2.404**	**0.018**
	Test Condition	16.578	0.473	0.637
Closeness Rating	**(Intercept)**	**262.565**	**7.006**	**< 0.001**
	Intelligibility Rating	**0.378**	**4.098**	**< 0.001**
	Test Condition	24.615	0.551	0.583

Table 2.7 Model parameters and statistics for regressions of Warmth, Competence, and Closeness rating over Intelligibility rating. Bold effects are significant at $p < 0.05$.

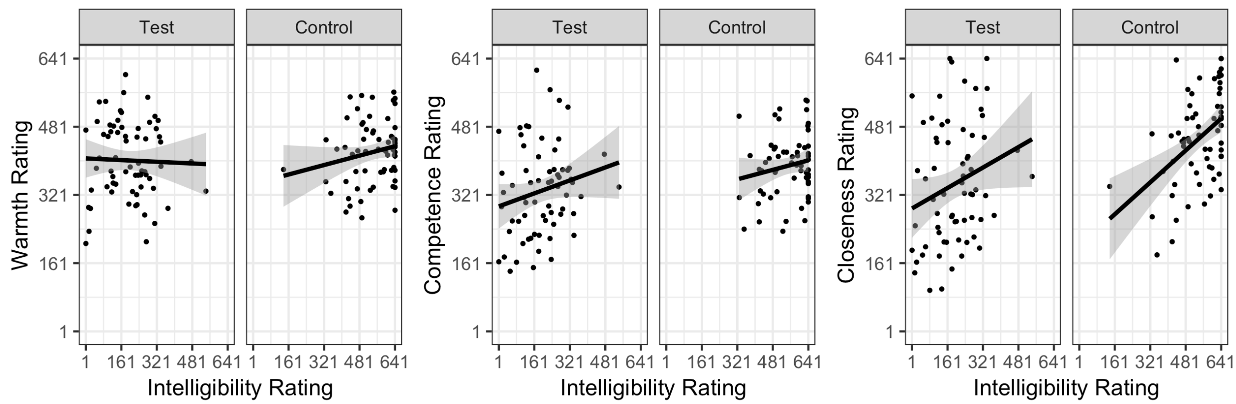

Figure 2.8 Warmth (right), Competence (middle), and Closeness (right) rating by Intelligibility rating across the Initial, Medial, and Final blocks. Intelligibility rating is significantly correlated with attitude rating on all three dimensions.

2.11.4 The Relationship Between Changes in Fluency and Attitude

Summarizing the results presented thus far, there is strong evidence that some measures of objective processing fluency improve over time as listeners gain experience with a non-native accent, but there was no direct evidence for an improvement in subjective processing fluency or attitudes (section 2.11.2). The analysis of the relationship between processing fluency measures and attitude rating summed across blocks found evidence suggesting that it is the subjective experience of ease, but not objective processing ease, that is related to positive attitudes (section 2.11.3).

Turning to the relationship between change in fluency and change in attitude over time, the prediction that increased processing fluency would positively correlate with improved attitudes was largely not upheld. For the analysis of change from the Initial to Medial block, no measure of change in fluency (Dilation, Edit Distance, Transcription Time, and Intelligibility rating) was significantly related to change in Warmth, Competence, or Closeness rating. The modeled results for change in Competence rating (summarized in Table 2.8) showed a marginally significant main effect of Condition ($\beta = 60.608$, $t = 1.892$, $p = 0.065$), indicating that

Competence rating increased from the Initial to Medial Block in the Test condition and decreased in the Control condition, consistent with the results for Competence ratings presented in section 2.11.2.

Dependent Variable	Fixed Effect (Predictor Variable)	β	t	p
Initial to Medial Competence Rating Change	(Intercept)	1.289	0.081	0.936
	Test Condition	*60.608*	*1.892*	*0.065*

Table 2.8 Model parameters and statistics for Initial-Medial Competence change. Italic effects are marginally significant at $p < 0.1$.

Modeled results comparing changes from the Medial to Final block were also largely inconsistent with predictions. None of the objective measures of change in fluency (Dilation, Edit Distance, and Transcription Time) was significantly related to change in Warmth, Competence, or Closeness rating. However, consistent with predictions, change in Intelligibility rating (Figure 2.9) had a marginally significant positive correlation with change in Competence rating ($\beta = 0.307$, $t = 1.717$, $p = 0.094$) and a significant positive correlation with change in Closeness rating ($\beta = 0.568$, $t = 3.348$, $p = 0.002$), but no significant correlation with change in Warmth rating. The effect of Condition and its interaction with Intelligibility change were also not significant for any of the three rating dimensions. Parameter estimates and statistics for these models are shown in Table 2.9.

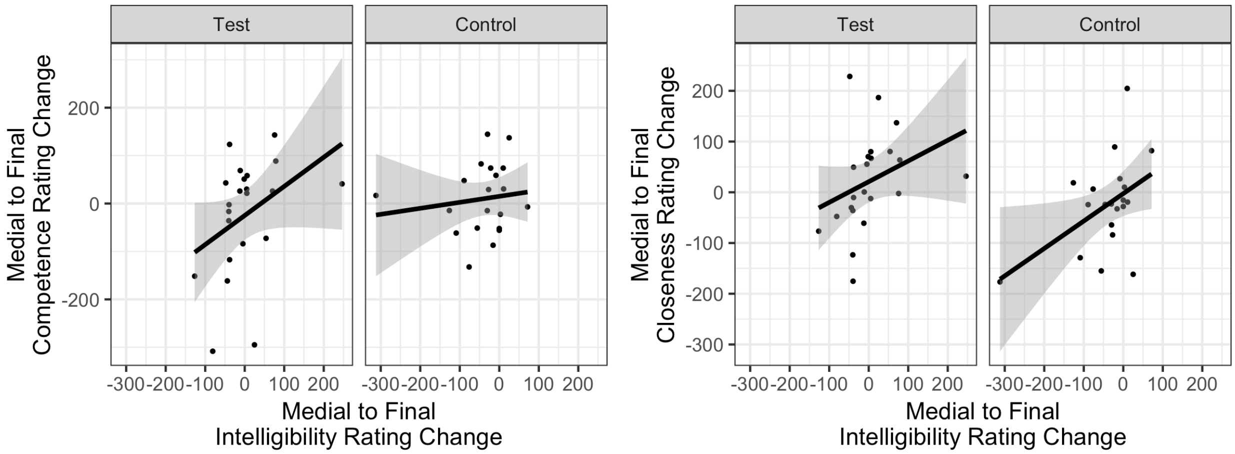

Figure 2.9 Medial-Final change in Competence and Closeness rating by change in Intelligibility rating. The effect of Intelligibility rating change was marginally significant for Competence rating change and significant for Closeness rating change.

Dependent Variable	Fixed Effect (Predictor Variable)	β	t	p
Medial to Final Warmth Rating Change	(Intercept)	-1.610	-0.126	0.901
	Intelligibility Rating Change	0.225	1.530	0.134
	Condition	-10.015	-0.389	0.700
Medial to Final Competence Rating Change	(Intercept)	-1.657	-0.107	0.916
	Intelligibility Rating Change	*0.307*	*1.717*	*0.094*
	Condition	-45.446	-1.453	0.154
Medial to Final Closeness Rating Change	(Intercept)	11.451	0.776	0.442
	Intelligibility Rating Change	**0.568**	**3.348**	**0.002**
	Condition	18.233	0.614	0.543

Table 2.9 Model parameters and statistics for Medial-Final attitude change and fluency change. Bold effects are significant at $p < 0.05$. Italic effects are marginally significant at $p < 0.1$.

2.11.5 The Relationship Between Fluency Measures

The physiological measure of processing fluency used in these experiments, pupil dilation, was predicted to positively correlate with the behavioral measures of transcription time and transcription errors. If these objective measures are not correlated, it may indicate that the increased comprehension accuracy of accented speech over time that has been observed in previous research was due to participants expending additional cognitive effort to understand an accent, rather than perception becoming easier through adaptation. Model estimates of the effects of the variable Dilation on the variables Transcription Time and Edit Distance in the Initial, Medial, and Final blocks (Figure 2.10) showed the expected positive relationship between the physiological and behavioral measures, consistent with the interpretation that accuracy and fluency increase over time as a result of adaptation. First- and second-order polynomial Dilation predictors had significant effects on Transcription Time ($Dilation^1$ $\beta = 0.486$, $t = 3.258$, $p = 0.001$; $Dilation^2$ $\beta = 0.778$, $t = 5.665$, $p < 0.001$), indicating a positive correlation for positive Dilation values but a negative correlation for negative Dilation values (Figure 2.10, left panels). Dilation also had a significant positive correlation with Edit Distance ($\beta = 0.290$, $z = 4.178$, $p < 0.001$; Figure 2.10, right panels). A summary of the parameter estimates and statistics for these models is shown in Table 2.10.

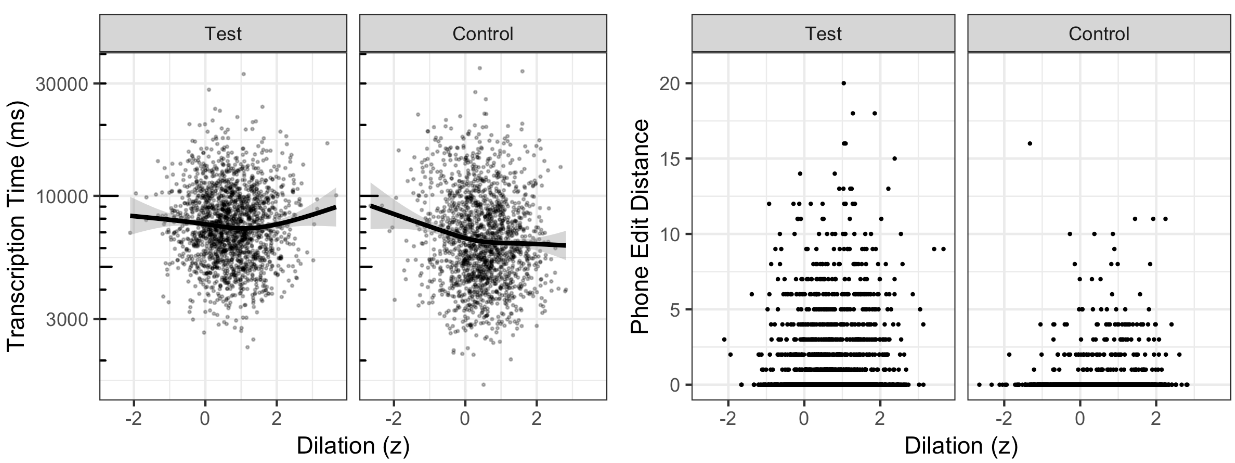

Figure 2.10 Transcription Time (left) and Edit Distance (right) by Dilation. Transcription Time is plotted on a $\log_{10}$-adjusted scale. The first- and second-order polynomial effects of Dilation on Transcription Time were both significantly positive. The plotted regression curve over Dilation was fit using a Generalized Additive Model to illustrate the polynomial relationship. The positive effect of Dilation on Edit Distance was significant, and Edit Distance was significantly higher in the Test condition than the Control condition.

Dependent Variable	Fixed Effect (Predictor Variable)	β	Test Statistic[a]	p
Transcription Time (log_{10} ms)	**(Intercept)**	**3.843**	**203.947**	**< 0.001**
	Dilation[1]	**0.486**	**3.258**	**0.001**
	Dilation[2]	**0.778**	**5.665**	**< 0.001**
	Test Condition	0.060	1.596	0.118
Edit Distance	**(Intercept)**	**-0.700**	**-3.948**	**< 0.001**
	Dilation	**0.290**	**4.178**	**< 0.001**
	Test Condition	**1.886**	**5.328**	**< 0.001**
	Dilation * Condition	-0.207	-1.490	0.136

Table 2.10 Model parameters and statistics for the effects of Dilation and Condition on Transcription Time and Edit Distance. Bold effects are significant at $p < .05$.
[a] For Edit Distance, the test statistic is *z*. For Transcription Time, the test statistic is *t*.

If participants' subjective experience of processing fluency reflects objective processing fluency, Intelligibility ratings should negatively correlate with Dilation, Edit Distance, and Transcription Time. The results presented in sections 2.11.3 and 2.11.4, however, indicate that subjective ease is correlated with positive attitudes, while objective ease is not. This suggests that Intelligibility ratings may reflect or be strongly mediated by social judgment. To test the effects of objective fluency versus attitude on subjective fluency, separate models were constructed for the three Intelligibility items (see section 2.4.1). If participants' perception of comprehension accuracy is related to objective accuracy (measured against the speaker's intended utterance), Edit Distance is predicted to negatively correlate with Transcribe item ratings ("I am able to accurately transcribe what this person says."). If participant's perception of processing ease is related to objective ease, Dilation is predicted to negatively correlate with Comprehend item ratings ("I can comprehend this person without difficulty."). The Speaks item ("This person speaks in a way that is easy to understand.") is somewhat ambiguous: if it is interpreted as a description of the listening experience, ratings are predicted to negatively correlate with Dilation;

if it is interpreted as a statement about an attribute of the speaker, ratings are predicted to positively correlate with attitude rating.

These predictions were tested using LMEMs for Transcribe, Speaks, and Comprehend ratings, pooling data from the initial, medial, and final blocks. Each model was constructed by adding the relevant objective fluency measure and Warmth, Competence, and Closeness rating as predictor variables one at a time to a base model with Condition as a fixed effect and a random intercept by Participant. Interactions of the predictor variables with Condition were also tested. Variables that did not significantly improve the model were omitted. Parameters and statistics for these models are shown in Table 2.11. Contrary to predictions, there were no significant effects of objective fluency measures on Intelligibility item ratings. However, Closeness rating had a significant positive relationship with Transcribe ($\beta = 0.301$, $t = 3.111$, $p = 0.002$), Speaks ($\beta = 0.321$, $t = 4.635$, $p < 0.001$), and Comprehend ($\beta = 0.229$, $t = 2.796$, $p = 0.006$) ratings. The interaction of Closeness and Condition was also significant for Speaks ratings ($\beta = -0.292$, $t = -2.110$, $p = 0.037$), indicating that Closeness had a more positive effect on Speaks ratings for the native guise than the non-native guise. Competence rating had a significant positive effect on Comprehend ratings ($\beta = 0.266$, $t = 2.779$, $p = 0.006$). There were no significant effects of Warmth rating on Intelligibility ratings. It should be noted that these tests conceptually duplicate the regressions of Warmth, Competence, and Closeness rating over Intelligibility rating (calculated as the mean of the Transcribe, Speaks, and Comprehend ratings) in section 2.11.3. However, the small *p*-values indicate that the results are significant even with multiple tests and provide evidence that Intelligibility and Closeness rating are positively related. There is no evidence for a relationship between Intelligibility ratings and objective fluency measures.

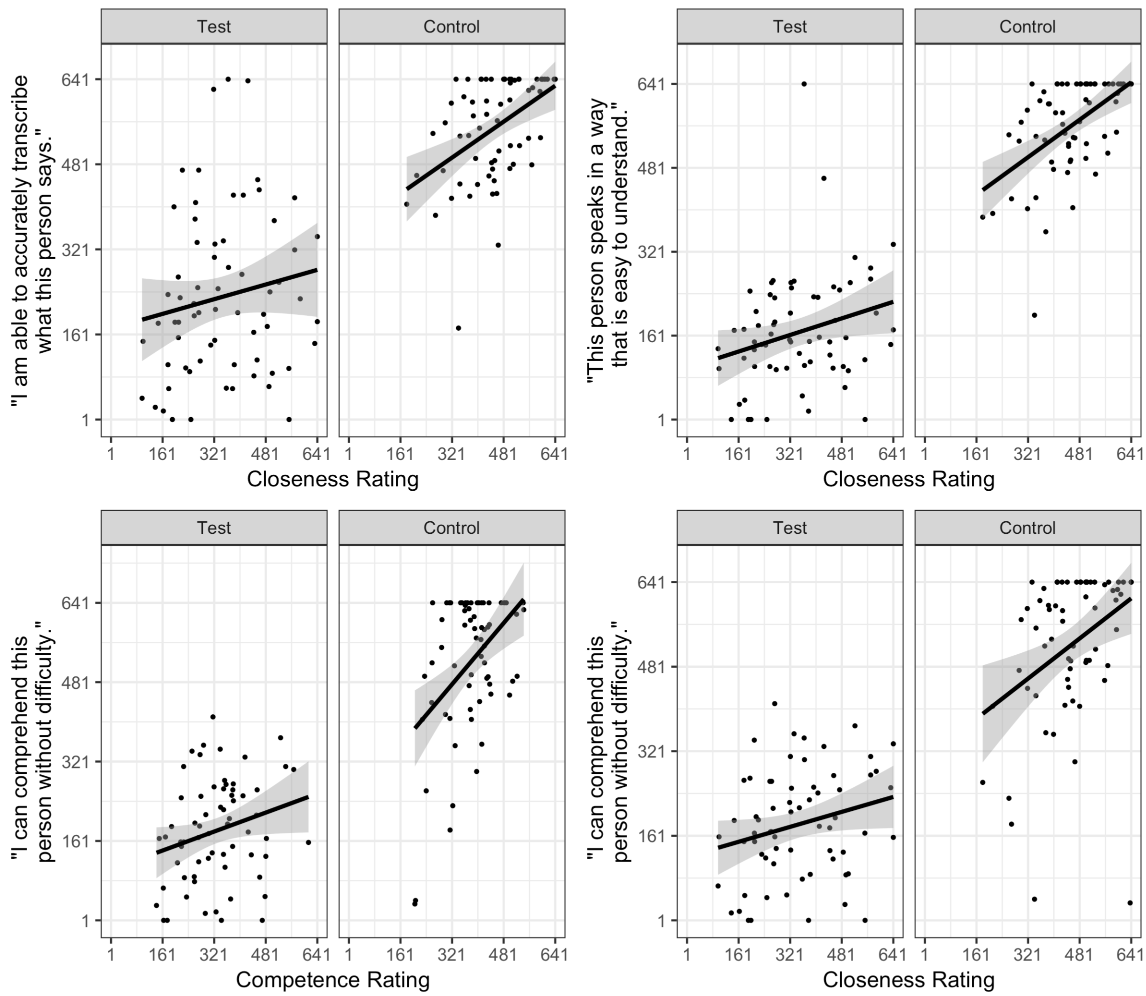

Figure 2.11 Top panels: Closeness rating by Transcribe (left) and Speaks (right) ratings. Bottom panels: Competence (left) and Closeness rating (right) by Comprehend ratings. All regression lines indicate significant effects, and the effect of Condition is significant for all dependent variables. For Speaks (top right), the interaction of Closeness and Condition is significant.

Dependent Variable	Fixed Effect (Predictor Variable)	β	t	p
Transcribe Rating	**(Intercept)**	**270.305**	**6.527**	**< 0.001**
	Closeness Rating	**0.301**	**3.111**	**0.002**
	Test Condition	**-285.912**	**-9.080**	**< 0.001**
Speaks Rating	**(Intercept)**	**227.487**	**7.337**	**< 0.001**
	Closeness Rating	**0.321**	**4.635**	**< 0.001**
	Test Condition	**-241.594**	**-3.896**	**< 0.001**
	Closeness * Condition	**-0.292**	**-2.109**	**0.037**
Comprehend Rating	**(Intercept)**	**165.648**	**3.543**	**0.001**
	Closeness Rating	**0.229**	**2.796**	**0.006**
	Competence Rating	**0.266**	**2.779**	**0.006**
	Test Condition	**-301.563**	**-9.527**	**< 0.001**

Table 2.11 Model parameters and statistics for the effects of Closeness and Competence rating and Condition on Intelligibility item ratings. Bold effects are significant at $p < 0.05$.

2.12 Discussion

The results from Experiment 1 provide clear evidence to support the hypothesis that objective processing fluency, as measured by both pupil dilation and transcription time, improves with experience with a non-native accent, but no evidence that transcription accuracy improves with experience. However, the significant positive relationship between dilation and transcription errors indicates that an increase in transcription accuracy (which has been reported in previous phonetic perceptual adaptation research) reflects greater ease, not greater effort, in comprehension. The absence of evidence for improved comprehension over trials may be the result of near-ceiling accuracy in both conditions. If this is the case, comprehension is expected to improve over time in Experiment 2, in which accuracy is expected to be lower for participants who hear sentence recordings mixed with noise.

Results measuring participants' subjective experience of processing fluency are less clear. Contrary to expectations, Intelligibility ratings decreased over time in the Control condition but did not change significantly in the Test condition. If the greater subjective processing difficulty in the Control condition is due to fatigue resulting from the experimental task, the unchanged Intelligibility ratings in the Test condition may indicate that improved subjective fluency due to perceptual adaptation counteracts the generally negative task effects for Test participants. However, it may also be that Intelligibility ratings near ceiling (see Figure 2.6) early in the experiment in the Control condition allow a greater range for decrease as participants gather more data about their ease and success in comprehending the speaker. That is, when subjective fluency is high early in the experiment, as it was for Control participants, a single transcription perceived as unsuccessful communication may reduce perceived ease. For Test participants, whose Intelligibility ratings did not change over time, gradual comprehension improvement over the experiment may not be recognized or not have an impact on the perception of disfluency that is established early in the experiment.

The evidence for increased fluency over time in these results is based on objective but not subjective measures: intelligibility ratings did not increase as pupil dilation or transcription time decreased. It is possible that subjective processing fluency is directly related to objective fluency measures in a way that this experiment could not identify using sentence-level objective measures (pupil dilation and edit distance) that were averaged over sentences for comparison with Intelligibility ratings that occurred only once per block. The relation between objective fluency and participants' evaluation of their own experience is also likely to be different depending on how much listening experience they reflect on. Participants in this experiment were instructed to rate the Intelligibility items based on their "most recent listening experience," so it is difficult to judge what information participants used in evaluating ease or difficulty. (However, it seems unlikely that participants could have precisely followed more specific instructions to rate Intelligibility based on, for example, the 10 most recent sentences they transcribed.) Reliable conclusions about the relationship between subjective and objective fluency may require subjective measures obtained more frequently relative to objective measures.

The correlation of subjective measures of fluency with judgments of social closeness, however, suggests that participants' self-reported experience of processing ease may reflect a social judgment rather than an appraisal of fluency. If this is the case, the low intelligibility ratings for the non-native guise may have been the result of stereotypes associated with non-native accents in general or identification of the speaker as a member of a different social group. Research in social psychology has shown that people require more information to classify another person as an in-group member than they do to classify a person as an out-group member (Leyens & Yzerbyt, 1992). If intelligibility ratings of the non-native guise were the results of social perception or stereotypes of the speaker, who was likely recognized almost immediately as an out-group member, it may be that increased fluency over time was insufficient to counter those judgments, and the stable ratings in the test condition may indicate the stability of listeners' social evaluation of the speaker. For the native guise, it may be that, as participants in this experiment gathered more information about the target speaker in the native guise over trials, they became less likely to identify him as an in-group member. The inclusion of sentence recordings mixed with noise in Experiment 2 may clarify how subjective processing fluency interacts with social judgments based on accent. Noise is expected to decrease objective processing fluency for the target speaker in the native guise. If the subjective experience of fluency is influenced more by social judgment than objective fluency, the effects of objective disfluency on intelligibility are predicted to be different for disfluency caused by noise versus a non-native accent.

The results of Experiment 1 also provide no evidence that speaker evaluation on the dimensions of warmth and competence improves with experience or that processing fluency and attitudes are related, though this may be due, at least in part, to methodology rather than a genuine absence of such effects. Measuring attitude change in this experiment depends on all participants interpreting adjectives in the same way. In the LMEMs for Warmth and Competence, the random effect estimates (indicating adjustments to the model intercept by Item) differed between adjective sets, indicating that the ratings from Bergsieker et al. (2012) may not have matched the favorability judgments of participants in this study. In Experiment 2, adjective favorability was rated by each participant, and those ratings indicate that many adjectives are interpreted as favorable by some participants and unfavorable by others. If this is the case for participants in Experiment 1, the reverse scoring of items based on the favorability ratings reported by Bergsieker and colleagues may have distorted the results, undermining their interpretability. The Warmth and Competence results reported in this chapter are likely not conclusive, pending the results of Experiment 2, in which variation in adjective favorability between participants was taken into account.

CHAPTER 3

Experiment 2

Experiment 1 provided evidence that both pupil dilation and transcription time decrease over transcription trials, indicating an improvement in objective processing fluency as listeners gain experience. However, the results provided no direct evidence of improvement in intelligibility ratings (subjective fluency) or attitude ratings over time. There was also no evidence of a relationship between objective fluency and either intelligibility rating or attitudes. These results may indicate that intelligibility and attitude ratings were not influenced by objective fluency, but it may also be that confounding factors obscured objective fluency's effects. First, the LMEM-estimated random effects of the adjective items used to test Warmth and Competence ratings in Experiment 1 indicated that participants' interpretation of the adjectives' favorability differed across the sets presented in different blocks of the experiment. (A different set of adjectives was used in each block to prevent participants' rating of a trait in one block influencing their rating of the same trait in other blocks.). Additionally, transcription errors were near floor for all participants, meaning participants may not have perceived communication to be hindered or may not have been aware of improvement over time.

To address these potential confounding factors, Experiment 2 adapted the earlier methodology by (1) presenting audio stimuli mixed with noise (in order to increase processing effort) for half of participants and (2) introducing a norming procedure for adjective favorability and speaker ratings. Additionally, the results of Experiment 1 indicated that pupil dilation and transcription errors are positively correlated, and pupil measurement was excluded from Experiment 2 to simplify the experimental procedure and decrease participant discomfort. These changes are designed to further test the main hypotheses that (1) processing fluency will increase with exposure to a non-native accent, in terms of both objective (i.e., comprehension accuracy and transcription time) and subjective (i.e. self-report) measures; and (2) speaker evaluation will improve with exposure to a non-native accent, in direct relation to the improvement in processing fluency. The added condition in which audio is presented in noise tests the corollary hypotheses that objective fluency, subjective fluency, and speaker evaluation will be negatively affected by noise and improve as participants adapt. This chapter first describes changes to the methods and materials used in Experiment 1, followed by predictions, results and statistical analysis, and a brief discussion of the results. A more comprehensive general discussion of the results of both experiments is presented in Chapter 4.

3.1 Participants

Data were collected from 76 undergraduate students at the University of Michigan, although data were analyzed for only 71 of these participants. Three participants reported native languages other than English in the language history questionnaire completed after the experiment (these were Tamil, Spanish, and Romanian), and their data were excluded from analysis. Two additional participants' mouse-tracking data showed that they did not move the mouse on more than 65% of 102 speaker rating items and simply clicked through the question. This was interpreted as evidence of distraction or lack of engagement with the experiment, and their data were also excluded. Of the remaining 71 participants, most failed to move the mouse on one or more rating items, but all of them moved the mouse on at least 70% of items. Of these

71 participants, 55 self-identified as women and 16 as men. All reported English as a native language and their dominant language; 13 reported an additional native language (these were Indonesian, Telugu, Romanian, Yoruba, Serbian, Gujarati [2 participants], Cantonese, Latvian, Marathi [2 participants], Hindi, and Punjabi). All participants reported normal hearing and normal or corrected vision. Participants were paid $10 for completing the experiment.

3.2 Materials

Half of the participants in Experiment 2 heard the same sentence recordings as were used in Experiment 1, while the other half heard the sentence recordings mixed with noise. The method used to create the noisy recordings is described in section 3.3. The presentation, rating, and scoring of the social perception (but not the intelligibility or social closeness) scale-response items in Experiment 2 differed from Experiment 1, and the new procedure is described in section 3.5.

Experiments 1 and 2 used the same sentence stimuli (see section 2.3), sentence ordering procedure (for predictable and unpredictable sentences; see section 2.3.1, final paragraph), scale-response items (see section 2.4), and language history and debriefing questionnaire (see section 2.5). Sentence stimuli are listed in Appendix A and the questionnaire is included in Appendix B. Intelligibility and social closeness scale-response items are repeated here for ease of reference. As in Experiment 1, each rating item was displayed separately on a screen with a continuous response scale labelled “Strongly Disagree” and “Strongly Agree” at the left and right endpoints, respectively. The three intelligibility items are:

1. I am able to accurately transcribe what this person says.
2. This person speaks in a way that is easy to understand.

3. I can comprehend this person without difficulty.

The three sets of social closeness statements are shown in Table 3.1.

Domain	Set 1	Set 2	Set 3
Working	I would work with this person on a group project.	I would work on a team with this person.	I would work together with this person on an assignment.
Socializing	I would get coffee with this person.	I would have lunch with this person.	I would play board games with this person.
Living	I would be happy to have this person as a neighbor.	I would be happy to live next door to this person.	I would be happy if this person moved in to the house next to mine.

Table 3.1 Social closeness rating statements.

3.3 Sentence Recordings Mixed with Noise

Speech-shaped noise was creating using a Praat script (Northwestern University Linguistics Labs, 2015) based on the long-term average spectrum of recordings for all sentence stimuli by the baseline speaker and the target speaker in both guises. All recordings for presentation were trimmed using SoX (Open Source Collaboration, 2015) to remove any amount of leading or trailing silence and normalized to 71 dB average intensity. 500 ms of silence was then appended to the beginning and end of each recording. Each file was mixed with speech-shaped noise using another Praat script (Northwestern University Linguistics Labs, 2015). Signal-to-noise ratios ranging from +0 dB (i.e., speech and noise of the same average intensity) to +5 dB (i.e., average speech intensity greater than average noise intensity) were tested with pilot participants to identify a ratio that made comprehension difficult but not frustrating, and a +3 dB ratio was selected for use in Experiment 2.

3.4 Social Perception and Adjective Rating

As in Experiment 1, participants' social perception of the speaker's warmth and competence attributes were measured by ratings of the accuracy of short adjective descriptions of the speaker (e.g., "This person is intelligent.") on a scale from "Not at all" to "Very". Experiments 1 and 2 used the same 39 adjectives (24 warmth, 15 competence) divided into three sets in order to present different adjective statements in the initial, medial, and final blocks. The order of adjectives within each set was the same as in Experiment 1, such that across the sets, item order was matched so that participants rated items with the same dimension in the same order. Within sets, the same item order was used for all participants to prevent confounding adjective spillover effects: for example, the word or description "ignorant" may be rated differently depending on whether it follows "arrogant" or "kind." The sets of attitude items are shown in Table 3.3. As in Experiment 1, the order of sets was partially counterbalanced across three groups of participants in both the Test and Control conditions (see Table 3.2).

3.4.1 Adjective Favorability and Baseline Rating

Experiment 2 introduced an adjective favorability rating task at the beginning of the experiment. Participants were asked to assume that the adjective was used to describe a person and to rate each adjective on a scale from "Very unfavorable" to "Very favorable". For adjectives with more than one meaning (e.g., "sensitive" might be interpreted as a favorable or unfavorable trait in different contexts), participants were instructed to choose one meaning and interpret the adjective in that same way throughout the experiment. The favorability rating was used to determine the reverse scoring of unfavorable adjectives. Experiment 2 also collected speaker ratings for all 39 adjectives in the Baseline block. (In Experiment 1, the Baseline block used same set of 13 adjectives as the Initial block.) Including all 39 adjectives made it possible to test the internal consistency of the 24 warmth and 15 competence adjectives and to compare mean favorability for the three sets. As in Experiment 1, different sets of 13 adjectives were used in the Initial, Medial, and Final blocks.

For both the favorability rating and Baseline speaker rating, the presentation order of the three sets of adjectives matched the order of sets used in the Initial, Medial, and Final blocks (see Table 3.2). For example, counterbalancing group 3 (in both conditions) rated set 3 in the Initial block, set 2 in the Medial block, and set 1 in the Final block. In the favorability and Baseline ratings for this group, participants rated set 3, set 2, and set 1 in that order. Between each set, the favorability and Baseline ratings were interrupted by a screen with a progress message (e.g., "You've completed 13 adjectives.") to mimic the separation of sets in the Initial, Medial, and Final block ratings.

3.5 Procedure and Data Collection

Participants were tested individually or in pairs in a sound-attenuated booth in the Phonetics Lab at the University of Michigan. The purpose of the study, as explained to participants, was to examine "how English speakers use adjectives." (The stated purpose of Experiment 1 was to examine "how listeners perceive different speakers.") After providing informed consent, participants were given oral instructions to familiarize them with the procedure. The oral instructions were similar to those for Experiment 1 (see section 2.7), except that participants were informed that the recordings they would hear may have background noise, and that if they had difficulty identifying a word they should type a dash in its place. This instruction was added to encourage participants, particularly those hearing sentences mixed with noise, to provide partial transcriptions of difficult sentences rather than a blank transcription, as well as to facilitate accuracy scoring. Instructions were also added for the adjective favorability task. Table 3.2 shows the presentation order and stimuli used for the complete experiment in the Test and Control conditions.

Block	Task	Test			Control		
		CB1	CB2	CB3	CB1	CB2	CB3
	Adjective Favorability Rating	*Item sets 1, 2, 3*	*Item sets 2, 3, 1*	*Item sets 3, 2, 1*	*Item sets 1, 2, 3*	*Item sets 2, 3, 1*	*Item sets 3, 2, 1*
Baseline	Transcription 10 sentences	Baseline speaker (American English accent)					
	Intelligibility	Intelligibility items					
	Social perception	*Item sets 1, 2, 3*	*Item sets 2, 3, 1*	*Item sets 3, 2, 1*	*Item sets 1, 2, 3*	*Item sets 2, 3, 1*	*Item sets 3, 2, 1*
	Social closeness	Set 1	Set 2	Set 3	Set 1	Set 2	Set 3
Initial	Transcription 10 sentences	Target speaker, Non-native guise			Target speaker, Native guise		
	Intelligibility	Intelligibility items					
	Social perception	Item set 1	Item set 2	Item set 3	Item set 1	Item set 2	Item set 3
	Social closeness						
Medial	Transcription 40 sentences	Target speaker, Non-native guise			Target speaker, Native guise		
	Intelligibility	Intelligibility items					
	Social perception	Item set 2	Item set 3	Item set 2	Item set 2	Item set 3	Item set 2
	Social closeness						
Break (90 seconds)							
Final	Transcription 40 sentences	Target speaker, Non-native guise			Target speaker, Native guise		
	Intelligibility	Intelligibility items					
	Social perception	Item set 3	Item set 1	Item set 1	Item set 3	Item set 1	Item set 1
	Social closeness						
Language history and debriefing questionnaire							

Table 3.2 Experiment 2 presentation order and material. “CB” columns indicate partial counterbalancing of set presentation order. Each participant in both the test and control conditions was randomly assigned to one of three counterbalancing groups. Cells in italic font are new or altered from Experiment 1.

The computer presentation of Experiment 2 was the same as Experiment 1 with two exceptions. First, participants were not required to look at the screen while listening to aural stimuli, as their pupils were not being measured. Second, the scale used for adjective favorability and speaker rating was 2000 pixels (rather than 640 pixels) wide because Experiment 2 was presented on different monitors at a different resolution than Experiment 1.[7]

3.6 Data Processing for Analysis

3.6.1 Transcription Errors and Transcription Time

Sentence transcriptions were checked for typographical errors, misspellings, and other errors by the experimenter following the same procedures as in Experiment 1 (section 2.8.2). In two of the trials, it was clear the participant had pressed "return" accidentally, intending to type an apostrophe, which resulted in an incomplete transcription. These trials were excluded from analysis. Errors were scored using phone edit distance, as in Experiment 1.

A preliminary analysis of transcription time indicated that the variable was not easily interpretable as an indicator of processing effort. Short transcription times were sometimes associated with high accuracy, but a large number of trials were transcribed quickly because the participant omitted most of the sentence. This was most common in the test and noise conditions. Participants were instructed that they could type a dash if they had difficulty understanding a word or part of a sentence, and this instruction may have increased the likelihood that participants would submit a partial transcription in a short time. Because no clear basis for

[7] Ratings were analyzed using scale, rather than standardized, units so that variation in use of the response scale among participants and for different items would be represented in the appropriate random-effect terms of the statistical models.

distinguishing whether short transcription times indicated easy or effortful processing could be identified, transcription time was not analyzed for Experiment 2. Edit Distance was used as the sole objective measure of processing fluency. (The results of Experiment 1 indicated that transcription errors are positively associated with pupil dilation.)

3.6.2 Scale-Response Item Scoring

Scale rating responses ranged from 1 to 2001, corresponding to the number of pixels in the scale on the screen. As in Experiment 1, the intelligibility and social closeness items were designed with positive valence and raw scores were used in the analyses. For social perception adjective ratings, reverse scoring was based on participants' individual favorability rating for each adjective.

A summary of favorability judgments for the 39 adjectives is shown in Table 3.3, which shows the number of participants who rated each adjective as favorable or unfavorable. (Ratings on which a participant clicked the scale without moving the mouse were discarded, so some adjectives' values do not sum to 71. Those participants' speaker ratings for that adjective were also omitted from analyses.) Some of the favorability judgments are implausible on visual inspection. For example, two participants rated "stupid" as a favorable description of a person, and the scale response for one of these participants was at the "Very favorable" extreme. Such ratings may be due to error or misapprehension, and it seems unlikely that a participant's rating of the speaker for a description like "This person is stupid." would consistently reflect a favorable interpretation and should be scored so that high values indicate positive evaluation. Inspection of the rating distributions showed that when the percentage of participants in the minority was 10% or less, most of the minority's favorability ratings were more than two standard deviations from the mean of all participants. This was not the case for adjectives with larger majorities, such as “happy-go-lucky” and “sensitive,” whose distributions were more neutral. Based on this pattern, adjectives with a minority of 10% or less were omitted from those participants’ speaker ratings for analysis. There were 25 participants for whom ratings on at least one adjective were omitted. Of these, there were 22 participants for whom three or fewer adjectives were omitted, one participant for whom 10 adjectives were omitted, and two participants for whom 14 adjectives were omitted. Removing data for the three participants who were often in the minority for favorability ratings did not change the patterns of results for warmth and competence ratings. Adjectives for which the percentage of participants in the favorability minority was larger than 10% were reverse-scored for those participants (if any) who rated the adjective as unfavorable.

Set 1			Set 2			Set 3		
Item	+	–	Item	+	–	Item	+	–
happy-go-lucky	61	8	sensitive	36	31	pleasure-loving	52	16
courteous	71	0	faithful	70	0	**passionate**	70	**(1)**
methodical	61	7	**practical**	65	**(4)**	persistent	60	9
kind	71	0	**honest**	70	**(1)**	generous	71	0
stubborn	**(6)**	65	**aggressive**	**(7)**	63	reserved	22	45
arrogant	**(3)**	68	**quarrelsome**	**(1)**	69	**rude**	**(2)**	69
ignorant	**(3)**	67	**lazy**	**(4)**	66	**stupid**	**(2)**	69
ambitious	66	**(4)**	industrious	53	11	**alert**	66	**(2)**
efficient	70	0	**sophisticated**	69	**(2)**	intelligent	70	0
cruel	**(3)**	68	**deceitful**	**(5)**	66	**treacherous**	**(4)**	67
argumentative	13	56	**quick-tempered**	**(5)**	66	**boastful**	**(6)**	64
shrewd	12	56	naive	10	58	frivolous	11	58
humorless	0	71	**conceited**	**(4)**	67	**revengeful**	**(2)**	69

Table 3.3 Number of participants who rated each adjective favorable (+) or unfavorable (–). Ratings in parentheses were in a minority smaller than 10% and those adjectives were excluded for those participants.

All participants heard the same American English-accented speaker in the Baseline block and evaluated him on all 39 adjectives. This allows a direct comparison of ratings across the three sets of adjectives in order to identify if any set elicits more positive or negative evaluations of the same (baseline) speaker than the other sets. This information is potentially important because differences in speaker evaluation across the sets may obscure changes in evaluation from the Initial to Medial and Medial to Final blocks, where the three counterbalancing groups of participants rated different adjective sets in each (non-Baseline) block. Additionally, in the Medial and Final blocks, two counterbalancing groups rated the same set because set order was only partially counterbalanced.

Speaker ratings in the Baseline block were compared across adjective sets to determine whether rating adjustment would clarify the interpretation of ratings in the Initial, Medial, and Final blocks. The effects of adjective set on baseline speaker ratings for Warmth and Competence (Figure 3.1) were tested in two LMEMs with Set as a fixed effect and a random intercept by Participant. (Models including a main effect of Audio [Clear vs. Noise] and its interaction with Set showed no significant effects of Audio or the interaction.) The Set variable contrasts used simple coding with Set 1 as the reference level, so that ratings in Set 2 and Set 3 were compared to ratings in Set 1, but ratings in Set 2 were not compared to Set 3. For Warmth ratings, there was a significant effect of Set 2 ($\beta = 232.674$, $t = 7.5841$, $p < 0.001$) and a marginally significant effect of Set 3 ($\beta = 58.259$, $t = 1.912$, $p = 0.056$). For Competence, there was a significant effect of Set 3 ($\beta = 92.563$, $t = 2.680$, $p = 0.008$). Model parameters and estimates are shown in Table 3.4. To control for rating differences among adjectives and adjective sets, models of Warmth and Competence ratings in the Initial, Medial, and Final blocks included the Baseline rating and its interaction with Condition as fixed effects. However, the pattern of results does not change when the effect of Baseline rating is omitted from the analysis.

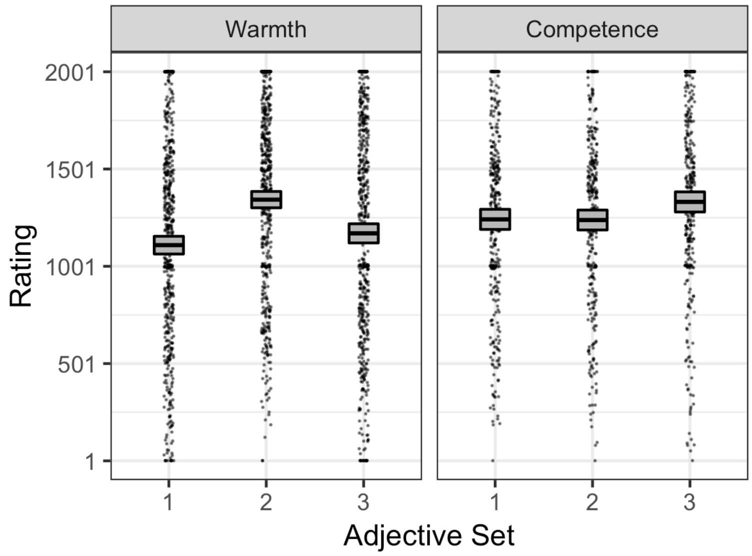

Figure 3.1 Baseline speaker ratings for adjectives by set. For warmth adjectives, Set 2 resulted in significantly higher ratings than Set 1. Set 3 ratings were marginally significantly higher than Set 1. For Competence ratings, Set 3 was rated significantly higher than Set 1.

Dependent Variable	Fixed Effect (Predictor Variable)	β	t	p
Warmth Rating	**(Intercept)**	**1206.959**	**51.131**	**< 0.001**
	Adjective Set 2	**232.674**	**7.584**	**< 0.001**
	Adjective Set 3	*58.259*	*1.912*	*0.056*
Competence Rating	**(Intercept)**	**1266.371**	**55.728**	**< 0.001**
	Adjective Set 2	-6.467	-0.187	0.852
	Adjective Set 3	**92.563**	**2.680**	**0.008**

Table 3.4 Model estimates and statistics for comparison of speaker rating by adjective set.

Ratings for Intelligibility, Warmth, Competence, and Closeness were pooled by dimension for analysis unless otherwise noted. As in Experiment 1, internal consistency across the three Intelligibility items was high in all four blocks (Baseline block $\alpha = 0.94$, Initial block α

= 0.97, Medial block α = 0.98, Final block α = 0.98). Internal consistency across Warmth and Competence items in the Baseline block was moderately high (Warmth α = 0.82, Competence α = 0.71). Calculating Cronbach's α is not possible for the Closeness items because the counterbalancing groups rated different items in each block. To account for variation across items in each dimension, a random intercept by Item was included in the respective models. (Full descriptions of the models are provided in sections 3.8.1–3.8.3.)

3.6.3 The Relationship Between Fluency and Attitude and Change over Time

As in Experiment 1, Edit Distance was adjusted by sentence length in phones and averaged over the last 10 trials in the initial, medial, and final blocks (i.e., immediately preceding the scale-response rating) to provide a summary statistic for comparison with Intelligibility, Warmth, Competence, and Closeness rating. To compare changes in Edit Distance and the rating variables over time, the difference of means between the initial and medial blocks and between the medial and final blocks were used as the change summary statistic.

3.7 Predictions

Experiment 2 tests the same predictions tested in Experiment 1. Objective fluency, as measured by Edit Distance, and subjective fluency, as measured by Intelligibility ratings, are predicted to improve over trials (i.e., Edit Distance is predicted to decrease and Intelligibility ratings are predicted to increase). Speaker evaluation, measured by Warmth, Competence, and Closeness ratings, is also predicted to improve (i.e., ratings are predicted to increase) from the Initial to Medial block and from the Medial to Final block. These changes are predicted to be larger in the Test condition than the Control condition. If participants are subjectively aware of increasing fluency over time, the means of Edit Distances over 10 transcription trials preceding Intelligibility rating are predicted to negatively correlate with Intelligibility item ratings. The hypothesized relationship between attitude and processing fluency predicts that mean Edit Distance for the 10 transcription trials preceding each rating will negatively correlate with Warmth, Competence, and Closeness rating. Intelligibility ratings are predicted to positively correlate with attitude rating.

Improvement in attitudes over time is hypothesized to depend on the improvement in objective and subjective processing fluency. Accordingly, the differences of means of Edit Distance between the Initial and Medial blocks and between the Medial and Final blocks are predicted to negatively correlate with differences in Warmth, Competence, and Closeness rating. The Initial-Medial and Medial-Final differences in Intelligibility rating are predicted to positively correlate with the corresponding differences in Warmth, Competence, and Closeness rating.

Based on previous research on the perception and evaluation of non-native speakers, Edit Distance is predicted to be higher on average in the Test condition than the Control condition. Intelligibility, Warmth, Competence, and Closeness ratings are predicted to be lower in the Test condition than the Control condition.

3.7.1 Predictions for the Effects of Noise

Noise is used in Experiment 2 for two main reasons: (1) to decrease comprehension accuracy, enabling better measurement of adaptation to the non-native accent over time, and (2) to decrease processing fluency for the native-accented guise in order to compare the effects of noise- and accent-driven disfluency on speaker perception. Accordingly, Edit Distances are predicted to be higher and Intelligibility ratings are predicted to be lower in Noise than Clear audio. The hypothesized relationship between fluency and attitudes also predicts that Warmth, Competence, and Closeness ratings will be lower for participants who hear recordings in Noise than Clear audio. Results will also be tested for an interaction of negative effects of noise and non-native accent on processing fluency. It is possible that noise-related disfluency will be attributed to the speaker differently depending on his accent. If disfluency caused by noise is attributed to the non-native accent but not the native accent, this will also result in interactive effects of noise and accent on speaker evaluation. Previous research has shown that listeners' comprehension accuracy of noise-masked and noise-vocoded speech (a different manipulation that also degrades acoustic cues to phonetic information) improves with experience (Banks et al., 2015; Huyck & Johnsrude, 2012; see also, Samuel & Kraljic, 2009), indicating that adaptation to noise may result in increased processing fluency. If this is the case, Noise and Condition (accent) are predicted to interact in their effects on change in transcription errors, intelligibility rating, and speaker evaluation over time.

3.8 Statistical Analysis

Predictions were tested using linear regression models as in Experiment 1. Edit Distance was modeled using a GLMEM with a Poisson-distributed dependent variable and a log link function, and test statistics were adjusted by the overdispersion to avoid anticonservative *p*-values. Necessary post-hoc tests were performed using the R *multcomp* package, with *p*-values adjusted using the single-step method. In all models that included the predictor variable Condition (Test vs. Control), the contrast values used simple coding with Control as the reference level. The predictor variable contrast for Audio (Noise vs. Clear) also used simple coding with Clear as the reference level. Models were first fit with full interactions between the predictor variable(s) of interest, Noise, and Condition. To reduce model overfitting for all variables except Edit Distance, higher-order interactions involving the predictor variable(s) of interest were removed when they did not significantly improve the model, as determined by an iterative process of nested-model comparison and reduction. Model comparison for Poisson GLMEMs with quasi-likelihood estimation is not currently implemented in the *lme4* package, and models of Edit Distance included two- and three-way interactions of the predictor of interest with Condition and Noise.

3.8.1 Processing Fluency and Attitudes for the Baseline Speaker in Noise vs. Clear Audio

Differences in Edit Distance, Intelligibility, Warmth, Competence, and Closeness ratings in the Baseline block between Noise and Clear audio were tested using models with Audio (Noise vs. Clear) as the fixed effect and a random intercept by Participant. The model of Edit Distance also includes a fixed effect of Sentence Length (in phones). Intelligibility, Warmth, Competence, and Closeness models include a random intercept by rating Item. The effects of Audio and its interaction with Condition on fluency and attitude ratings were also tested in the models described below.

3.8.2 Processing Fluency and Attitudes for Native vs. Non-Native Accents

Differences in Edit Distance, Intelligibility, Warmth, Competence, and Closeness ratings between the Baseline and Initial blocks were tested using models with Block (Baseline vs. Initial), Condition (Test vs. Control), Audio (Noise vs. Clear) and their interaction as fixed effects, with a random intercept by Participant. The model of Edit Distance includes a fixed effect of Sentence Length (in phones). Intelligibility, Warmth, Competence, and Closeness models include a random intercept by rating Item. The Block variable contrast uses simple coding with Baseline as the reference level. The main effects and interaction of Condition and Audio in the models described in section 3.8.2 also tested differences in fluency and attitude depending on accent and noise.

3.8.3 Change in Fluency and Attitude with Experience

Change in Edit Distance over trials in the Initial, Medial, and Final blocks was tested in a GLMEM with Trial Order (an integer ranging from 1 to 90), Condition (Test vs. Control), Audio (Noise vs. Clear), Sentence Length (in phones), and the interactions of Trial Order, Condition, and Audio as fixed effects. Differences in scale-response ratings between the Initial, Medial, and Final blocks were tested using models with Block (Initial vs. Medial, Medial vs. Final), Condition (Test vs. Control), Audio (Noise vs. Clear), and their interaction as fixed effects. Models of Warmth and Competence ratings also included a main effect of Baseline rating and its interaction with Condition. The Block variable contrast used backward difference coding to compare the Medial and Initial blocks and the Final and Medial blocks. In this coding system, the Final and Initial blocks are not directly compared. All models included a random intercept by Participant. Models with Intelligibility, Warmth, Competence, or Closeness as the dependent variable also included a random intercept by Item.

3.8.4 The Relationship Between Fluency and Attitudes

The effect of processing fluency on Warmth, Competence, and Closeness was tested in LMEMs using data from the Initial, Medial, and Final blocks together. Separate models were used for each attitude dimension. Each model was constructed by adding the predictor variables mean Edit Distance and Intelligibility rating one at a time as fixed effects to a base model with effects of Condition, Noise, their interaction, and a random intercept by Participant. ANOVA model comparisons were used to test whether the addition of each fluency variable significantly improved the model fit, and predictor variables that did not were omitted from the final model. Interactions of significant predictor variables with Condition (Test vs. Control) and Audio (Noise vs. Clear) were also added and tested.

3.8.5 Change in Fluency and Attitudes

The relationship between changes in processing fluency and changes in evaluation was tested in linear models with one data point per participant and did not include random effects. Six models were constructed altogether for Initial-Medial and Medial-Final rating differences for Warmth, Competence, and Closeness. Each model was constructed by the same iterative fixed-effect addition procedure described in section 3.8.4 using differences in mean Edit Distance and Intelligibility rating as predictor variables.

3.8.6 The Relationship Between Subjective and Objective Fluency Measures

As in Experiment 1, the relative effects of Edit Distance (measuring objective fluency) and attitude rating on subjective fluency were tested in separate models for Transcribe, Speaks, and Comprehend ratings (see section 3.2), pooling data from the initial, medial, and final blocks. Models testing these effects were constructed using the following procedure. First, a model was constructed with a three-way interaction of Edit Distance with Condition (Test vs. Control) and Audio (Noise vs. Clear) and a random intercept by Participant. Model comparisons indicated that the interactions were only significant when data were included from a participant in the Control × Clear condition who had exceptionally high mean Edit Distance and low Intelligibility ratings in the medial block, but not the initial or final blocks. Some of the participant's transcriptions in the medial block (but not the initial or final blocks) were truncated, indicating that this individual omitted the last few words of the sentence. None of these sentences was extremely long or unpredictable, and no explanation for the pattern could be found in the participant's responses to the debriefing questions. Because interactions of Edit Distance with Condition and Audio appeared to be driven by this participant alone, that individual's data were excluded and only a main effect of Edit Distance and a two-way interaction of Condition and Audio were retained in the models. Warmth, Competence, and Closeness rating were added one at a time as predictor effects to this base model, and variables that did not significantly improve the model were omitted. A three-way interaction of Closeness rating with Condition and Audio was significant in models for all three Intelligibility items. The main effects of Warmth rating, Competence rating, and their interactions with Condition and Audio were not significant in any model.

3.9 Results

3.9.1 Processing Fluency and Attitudes for Clear vs. Noise-Masked Speech

Processing was predicted to be less fluent for recordings mixed with noise. Consistent with this prediction, Edit Distances in the Baseline block (in which all participants heard the same native-accented American English speaker; Figure 3.2, left panel) were significantly higher in Noise than in Clear audio ($\beta = 1.925$, $t = 4.377$, $p < 0.001$), while Intelligibility ratings in the Baseline block (Figure 3.2, right panel) were significantly lower in Noise than in Clear audio ($\beta = -596.74$, $t = -6.588$, $p < 0.001$). Speaker evaluation was also predicted to be less positive when speech was mixed with noise. This prediction was not upheld: there were no significant effects of Audio on Warmth, Competence, or Closeness ratings in the Baseline block (Figure 3.3). Model estimates and statistics are shown in Table 3.5

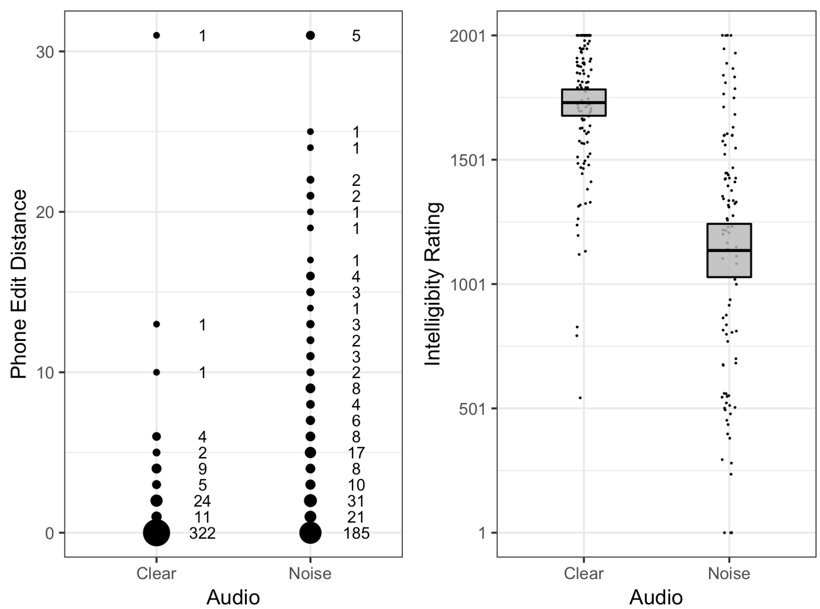

Figure 3.2 Edit Distances and Intelligibility ratings for speech in the Clear and Noise conditions in the Baseline block. The effect of Noise was significant for both dependent variables.

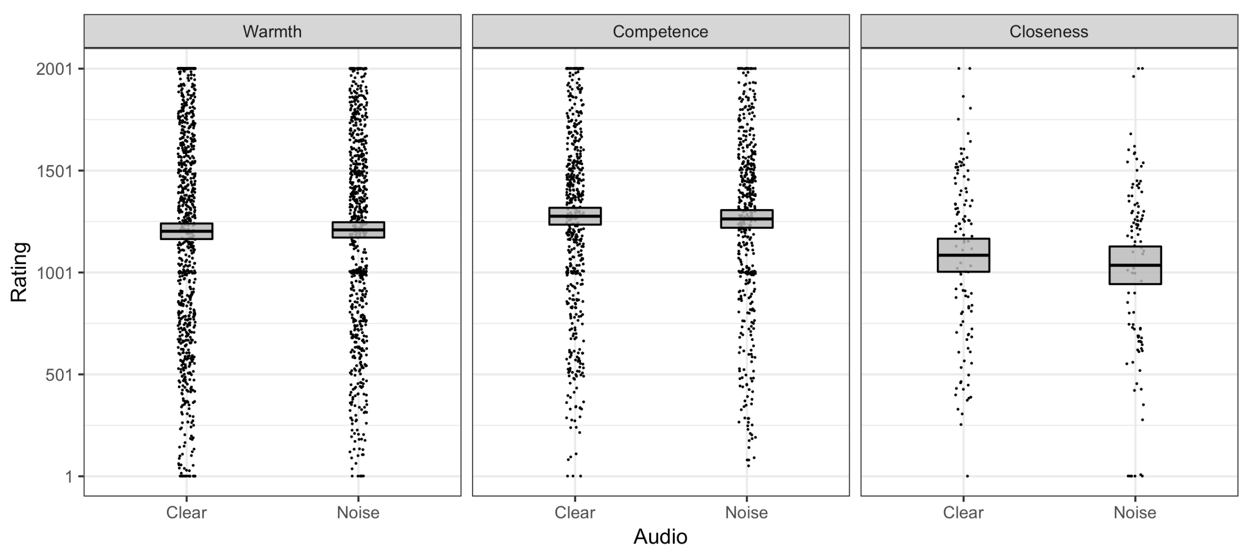

Figure 3.3 Warmth, Competence, and Closeness ratings for speech in the Clear and Noise conditions in the Baseline block. The effect of Noise on ratings was not significant.

Dependent Variable	Fixed Effect (Predictor Variable)	β	Test Statistic[a]	p
Edit Distance	**(Intercept)**	**-1.699**	**-4.582**	**< 0.001**
	Audio Noise	**1.925**	**4.377**	**< 0.001**
	Sentence Length (phones)	**0.053**	**6.280**	**< 0.001**
Intelligibility Rating	**(Intercept)**	**1432.473**	**26.742**	**< 0.001**
	Audio Noise	**-596.739**	**-6.588**	**< 0.001**
Warmth Rating	**(Intercept)**	**1208.426**	**17.865**	**< 0.001**
	Audio Noise	12.231	0.269	0.789
Competence Rating	**(Intercept)**	**1260.057**	**22.901**	**< 0.001**
	Audio Noise	-15.766	-0.345	0.731
Closeness Rating	**(Intercept)**	**1060.056**	**18.697**	**< 0.001**
	Audio Noise	-57.549	-0.718	0.475

Table 3.5 Model estimates and statistics for the effects of Noise vs. Clear audio in the Baseline block. Bold effects are significant at $p < .05$.
[a] For Edit Distance, the test statistic was *z*. For all other models, the test statistic was *t*.

3.9.2 Processing Fluency and Attitudes for Native vs. Non-Native Accents

Processing was predicted to be more fluent and speaker evaluation more positive for native-accented speech (i.e., in the Baseline block for all participants and the Initial block for Control participants) than non-native accented speech (i.e., in the Initial block for Test participants). For Edit Distance (Figure 3.4, left panels), the significant main effect of Initial Block and the significant interaction between Condition and Block indicate that errors increased from the Baseline to Initial block more in the Test condition than in the Control, reflecting less fluent processing of the non-native accent than the native accent, as predicted. The prediction was also upheld for Intelligibility (Figure 3.4, right panels), with significant effects of Block and the interaction of Condition and Block indicating that ratings decreased from the Baseline to Initial block more in the Test condition than the Control condition. Model parameters and statistics for the Baseline-Initial comparison of processing fluency are shown in Table 3.6.

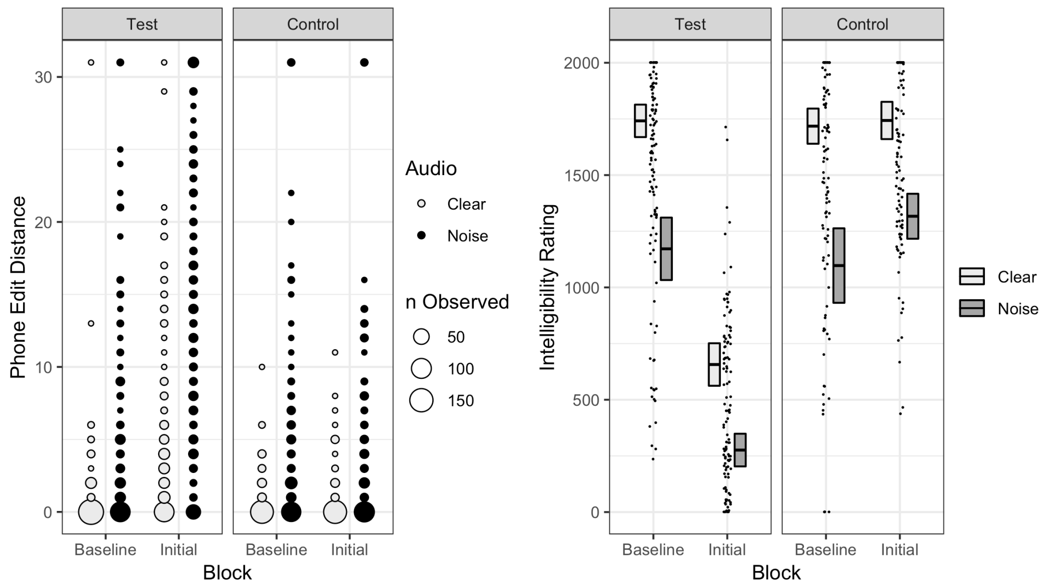

Figure 3.4 Edit Distance (left), and Intelligibility ratings (right) in the Baseline and Initial blocks. Cross bars for Intelligibility ratings indicate the mean and 95% confidence intervals. The interaction of Block and Condition was significant for both dependent variables

Dependent Variable	Fixed Effect (Predictor Variable)	β	Test Statistic[a]	p
Edit Distance	**(Intercept)**	**-0.581**	**-2.465**	**0.014**
	Initial Block	**0.730**	**5.686**	**< 0.001**
	Test Condition	**0.981**	**2.889**	**0.004**
	Noise Audio	**1.650**	**4.861**	**< 0.001**
	Sentence Length (phones)	**0.034**	**7.027**	**< 0.001**
	Block * Condition	**1.355**	**5.277**	**< 0.001**
	Block * Audio	-0.288	-1.123	0.261
	Condition * Audio	-0.288	-0.424	0.671
	Block * Condition * Audio	-0.156	-0.304	0.761
Intelligibility Rating	**(Intercept)**	**1214.906**	**38.097**	**< 0.001**
	Initial Block	**-433.384**	**-15.283**	**< 0.001**
	Test Condition	**-506.758**	**-7.945**	**< 0.001**
	Noise Audio	**-499.781**	**-7.836**	**< 0.001**
	Block * Condition	**-1113.206**	**-19.666**	**< 0.001**
	Block * Audio	**192.088**	**3.394**	**0.001**
	Condition * Audio	49.101	0.385	0.702

Table 3.6 Model estimates and statistics for the Baseline-Initial comparison of Edit Distance and Intelligibility ratings. Bold effects are significant at $p < 0.05$.

[a] For Edit Distance, the test statistic was *z*. For Intelligibility Rating, the test statistic was *t*.

Warmth, Competence, and Closeness ratings in the Baseline and Initial blocks are shown in Figure 3.5 and model estimates and statistics for the Baseline-Initial comparison are shown in Table 3.7. As in Experiment 1, Competence ratings (Figure 3.5, top-right panels) were consistent with the prediction that speaker evaluation would be more positive for native-accented speech than non-native accented speech. The significant effect of Block and the interaction of Condition and Block indicate that ratings decreased from the Baseline to Initial block more in the Test condition than the Control condition. For Closeness (bottom panels), the significant main effect of Block and the significant interaction of Condition and Block indicate that ratings increased from Baseline to Initial block in the Control condition but not the Test condition. This pattern, while not consistent with predictions, is the same as the pattern for Closeness ratings in the Baseline and Initial blocks in Experiment 1. Thus, Experiment 2 results further support the interpretation that Closeness ratings depend in part on a factor other than accent, such as voice quality, and participants preferred the target speaker to the baseline speaker. If this is the case, this preference may have been counteracted in the Test condition by the non-native accent.

Warmth ratings were not consistent with predictions, but were partly consistent with the results of Experiment 1. The significant effect of Block and the significant three-way interaction of Block, Condition, and Audio reflect the following pattern: Warmth ratings increased from Baseline to Initial block for both Test and Control conditions, but while ratings in the Control condition increased more in Noise than in Clear audio, ratings in the Test condition increased more in Clear audio than in Noise. Post-hoc tests indicated that ratings increased significantly in the Test × Clear ($\beta = 106.815$, $t = 2.5303$, $p = 0.013$), Control × Clear ($\beta = 126.613$, $t = 2.879$, $p = 0.005$), and Control × Noise ($\beta = 270.211$, $t = 5.967$, $p < 0.001$) conditions, but there was no significant difference between Baseline and Initial blocks in the Test × Noise condition. These patterns are illustrated by the raw data means shown in Figure 3.5 (top-left panels). The different effects of noise on evaluations of the target speaker (relative to the baseline speaker) in the non-native (Test) and native (Control) guises suggest that the disfluency caused by noise may have been attributed to the non-native accent but not the native accent. This possibility will be further discussed in section 3.10, in connection with other results.

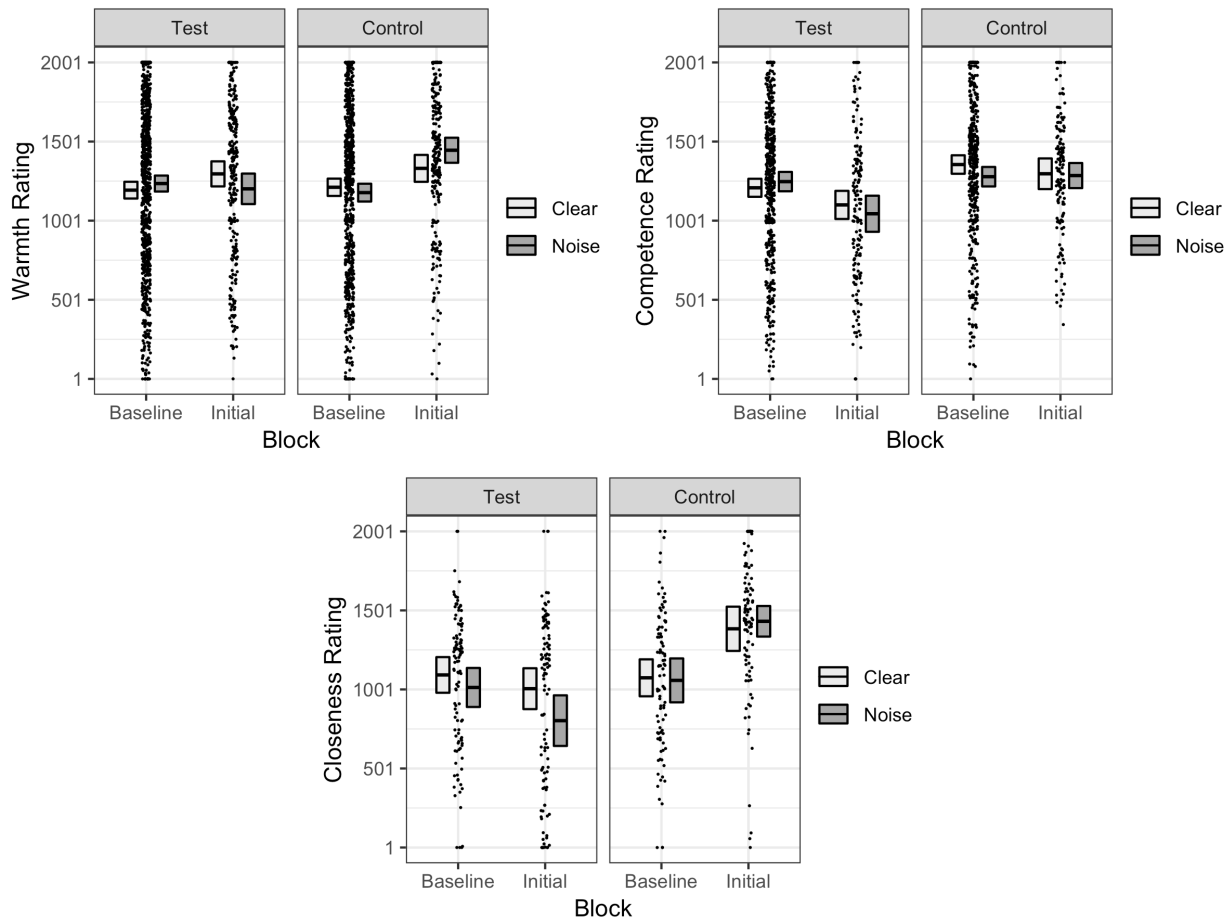

Figure 3.5 Warmth (top left), Competence (top right), and Closeness (bottom) ratings in the Baseline and Initial blocks. (See Table 3.1 and Table 3.3 for lists of items.) Cross bars indicate the mean and 95% confidence intervals. The effect of Block and the interaction of Block and Condition were significant for Competence and Closeness. The effect of Block and the interaction of Block, Condition, and Audio were significant for Warmth.

That noise significantly increased transcription errors and significantly decreased intelligibility ratings is consistent with the prediction that noise reduces both objective and subjective processing fluency. Although the absence of a significant effect of Audio on Warmth, Competence, and Closeness ratings (section 3.9.1) is inconsistent with the prediction that the decreased fluency caused by noise also generally worsens speaker evaluation, the three-way interaction of Block with Condition and Audio on Warmth ratings may indicate that noise has a negative effect on fluency and speaker evaluation (at least on the warmth dimension) for non-native accented speech but not native accented speech. These patterns will be revisited in sections 3.9.3 and 3.9.4 in connection with the analysis of fluency and attitudes in the Initial, Medial, and Final blocks.

Dependent Variable	Fixed Effect (Predictor Variable)	β	t	p
Warmth Rating	**(Intercept)**	**1263.409**	**59.272**	**< 0.001**
	Initial Block	**116.714**	**4.594**	**< 0.001**
	Test Condition	-58.208	-1.365	0.176
	Noise Audio	4.327	0.101	0.919
	Block * Condition	**-163.396**	**-3.216**	**0.001**
	Block * Audio	0.004	< 0.001	1.000
	Condition * Audio	-58.695	-0.688	0.493
	Block * Condition * Audio	**-287.195**	**-2.826**	**0.005**
Competence Rating	**(Intercept)**	**1227.124**	**57.652**	**< 0.001**
	Initial Block	**-85.389**	**-3.037**	**0.002**
	Test Condition	**-156.498**	**-3.676**	**< 0.001**
	Noise Audio	-27.376	-0.681	0.498
	Block * Condition	**-136.920**	**-2.435**	**0.015**
	Condition * Audio	76.224	0.949	0.346
Closeness Rating	**(Intercept)**	**1106.777**	**29.876**	**< 0.001**
	Initial Block	**103.200**	**2.731**	**0.007**
	Test Condition	**-259.390**	**-3.501**	**0.001**
	Noise Audio	-61.083	-0.824	0.413
	Block * Condition	**-479.225**	**-6.341**	**< 0.001**
	Condition * Audio	-156.736	-1.058	0.294

Table 3.7 Model parameters and statistics comparing Warmth, Competence, and Closeness ratings in the Baseline and Initial blocks. Bold effects are significant at $p < .05$.

3.9.3 Change in Fluency with Experience

Processing fluency, as measured by transcription errors and intelligibility ratings, was predicted to improve over time, with a larger improvement in the test condition (compared to the control condition) as participants gained experience with the non-native accent. Results for Edit Distance (Figure 3.6) are inconsistent with this prediction. Edit Distance decreased significantly over trials ($\beta = -0.005$, $z = -4.364$, $p < 0.001$) but the interaction between Condition and Trial Order was not significant. As in Experiment 1, transcription accuracy was at or near ceiling in a large majority of Clear Audio trials, but this was not the case for Noise Audio trials. Table 3.8 summarizes the percentage of trials in which phone edit distance was 0 (indicating no difference between the recorded sentence and the transcription) in the different conditions. The significant main effect of Condition in the model of Edit Distance across all trials ($\beta = 1.690$, $z = 6.165$, $p < 0.001$) supports the prediction that processing is more fluent for native than non-native accented speech, and the significant main effect of Audio ($\beta = 1.493$, $z = 5.451$, $p < 0.001$) supports the prediction that processing is more fluent for speech in clear audio than speech masked with noise. Parameter estimates and statistics for the model are shown in Table 3.9.

	Test Condition	Control Condition
Audio Clear	60.6%	88.3%
Audio Noise	27.4%	65.2%

Table 3.8 Percentage of trials with no transcription errors in the four sentence transcription conditions

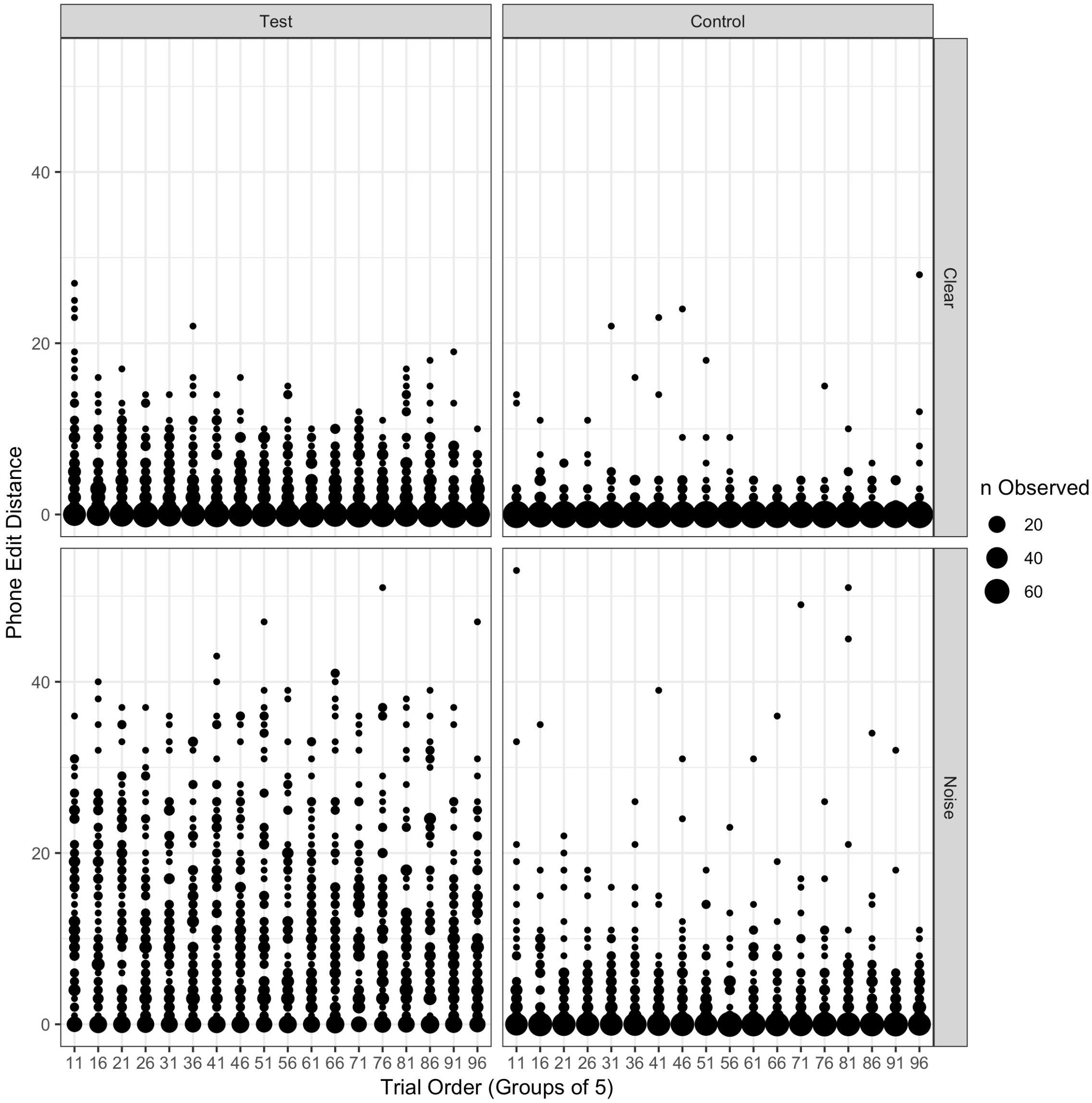

Figure 3.6 Edit Distances by Trial Order in the Initial (trials 11-20), Medial (trials 21-60), and Final (trials 61-100) blocks. The main effects of Trial Order, Condition, and Audio were significant.

Experiential effects on Intelligibility ratings (Figure 3.7) also differed from predictions. Intelligibility ratings were predicted to increase from Initial to Medial and from Medial to Final blocks, with a greater increase in the Test condition than the Control condition. In the model comparing ratings by Condition and Block, mean ratings decreased significantly from the Initial to Medial block ($\beta = -73.288$, $t = -3.113$, $p = 0.002$), and this effect showed a significant interaction with Condition (β = 186.221, t = 3.956, p < 0.001), indicating that the decrease was greater in the Control condition than the Test condition. Post-hoc tests indicated that ratings decreased significantly in the Control condition (β = -166.197, t = -4.855, p < 0.001) but did not change significantly in the Test condition. There were no significant effects or interactions for the comparison from the Medial to Final block. (A decrease in ratings for the Control condition but not the Test condition was also observed in Experiment 1, but in Experiment 1 the decrease occurred from the Medial to Final block.) The significant main effects of Condition (β = -943.349, t = -16.132, p < 0.001) and Audio (β = -389.543, t = -6.662, p < 0.001) support the predictions that participants would perceive processing to be easier both for a native accent than a non-native accent and for speech masked with noise than speech in clear audio. However, the interaction between Condition and Audio was not significant, indicating that the effect of noise on subjective perception of processing ease does not interact with the effect of accent. (Parameter estimates and statistics from this LMEM are shown in Table 3.9.)

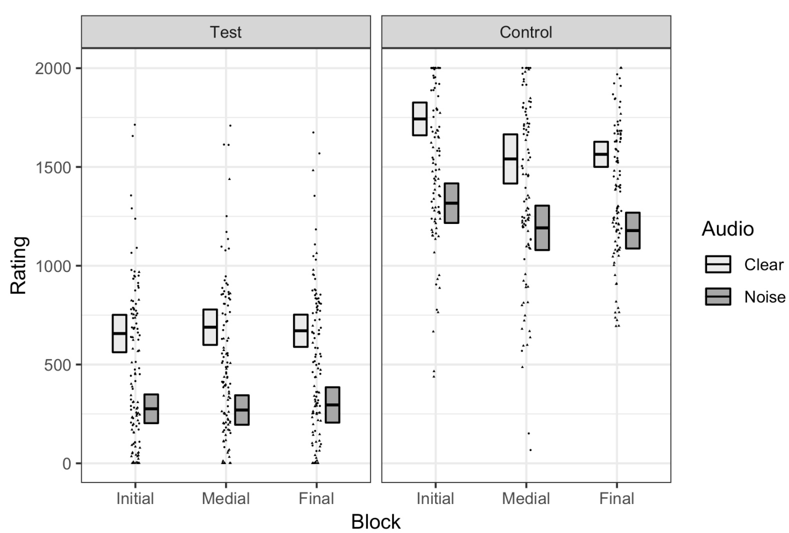

Figure 3.7 Intelligibility ratings by Block, Audio, and Condition. Ratings decreased significantly from the Initial to Medial block in the Control but not the Test condition. No significant differences were found between Medial and Final blocks.

Dependent Variable	Fixed Effect (Predictor Variable)	β	Test Statistic[a]	p
Edit Distance	**(Intercept)**	**-0.425**	**-2.782**	**0.005**
	Trial Order	**-0.005**	**-4.364**	**< 0.001**
	Test Condition	**1.690**	**6.165**	**< 0.001**
	Noise Audio	**1.493**	**5.451**	**< 0.001**
	Sentence Length (phones)	**0.039**	**19.426**	**< 0.001**
	Trial Order * Condition	-3.9×10^{-4}	-0.173	0.863
	Trial Order * Audio	0.003	1.359	0.174
	Condition * Audio	-0.270	-0.493	0.622
	Trial Order * Condition * Audio	0.005	1.032	0.302
Intelligibility Rating	**(Intercept)**	**949.855**	**28.481**	**< 0.001**
	Block 1 Initial to Medial	**-73.072**	**-3.120**	**0.002**
	Block 2 Medial to Final	1.028	0.044	0.965
	Test Condition	**-943.410**	**-16.132**	**< 0.001**
	Audio Noise	**-389.395**	**-6.658**	**< 0.001**
	Block 1 * Condition	**186.251**	**3.976**	**< 0.001**
	Block 2 * Condition	-9.128	-0.195	0.846
	Condition * Audio	-1.398	-0.012	0.990

Table 3.9 Model parameters and statistics for comparisons across trials/blocks. Bold effects are significant at $p < 0.05$.
[a] For Edit Distance, the test statistic was *z*. For Intelligibility Rating , the test statistic was *t*.

3.9.4 Change in Attitude with Experience

Warmth, Competence, and Closeness ratings in the Initial, Medial, and Final blocks are shown in Figure 3.8. Parameter estimates and statistics for the corresponding models are summarized in Table 3.10. Attitudes towards the speaker on each dimension were predicted to improve from Initial to Medial block and from Medial to Final block, with greater improvement in the Test condition than the Control condition. The results for Warmth (Figure 3.8, top-left panels) were not consistent with this prediction. There was a marginally significant main effect of Initial to Medial Block ($\beta = -40.956$, $t = -1.722$, $p = 0.085$) and a significant interaction of Initial-Medial Block with Condition ($\beta = 173.962$, $t = 3.912$, $p < 0.001$). Post-hoc tests indicate that Warmth ratings decreased significantly in the Control condition ($\beta = -127.937$, $t = -3.832$, $p < 0.001$) but did not change significantly in the Test condition. There was also a marginally significant interaction of Condition and Audio ($\beta = -227.542$, $t = -1.908$, $p = 0.061$). Post-hoc tests indicated that the difference between Noise and Clear audio was not significant in either the Test or Control condition; however, ratings were significantly lower in the Test condition than the Control condition in Noise ($\beta = -374.176$, $t = -3.814$, $p < 0.001$), but there was no significant difference between conditions in Clear audio. This is consistent with the increase in Warmth ratings from the Baseline to Initial Block for the Control condition but not the Test condition in Noise (see section 3.9.2).

Competence ratings (Figure 3.8, top-right panels) were largely inconsistent with predictions. The main effect of Condition was marginally significant ($\beta = -171.869$, $t = -1.869$, $p = 0.062$), supporting the prediction that ratings would be lower for the non-native guise than the native guise. However, while the effect of Initial to Medial Block was also marginally significant ($\beta = -58.733$, $t = -1.845$, $p = 0.065$), there were no significant interactions of Block and Condition, indicating that ratings decreased over time in both the Test and Control conditions, contrary to predictions. There was also a marginally significant interaction of Condition and Audio ($\beta = -206.293$, $t = -1.859$, $p = 0.067$). Post-hoc tests indicate that Noise had a negative effect on Competence ratings in the Test condition ($\beta = -163.644$, $t = -2.149$, $p = 0.035$) but there was no significant difference between ratings in Noise and Clear audio in the Control condition.

Closeness ratings (Figure 3.8, bottom panels) were also not consistent with the prediction that attitudes toward the speaker would improve over time more in the Test condition than the

Control condition. Contrary to predictions, there was a significant decrease in Closeness ratings from the Initial to Medial block ($\beta = -70.87$, $t = -2.104$, $p = 0.036$), but no significant change from the Medial to Final block and no significant interactions of Block with Condition or Audio. There was a significant main effect of Condition ($\beta = -445.04$, $t = -5.498$, $p < 0.001$), supporting the prediction that ratings would be lower for the non-native (Test) guise than the native (Control) guise. There was also a significant interaction of Condition with Audio ($\beta = -378.34$, $t = -2.337$, $p = 0.022$). Post-hoc tests indicate that Noise had a significant negative effect on Closeness ratings in the Test condition ($\beta = -316.383$, $t = -2.858$, $p = 0.006$) but there was no significant difference between Noise and Clear audio in the Control condition, consistent with the interaction effect of Condition and Audio on Competence ratings.

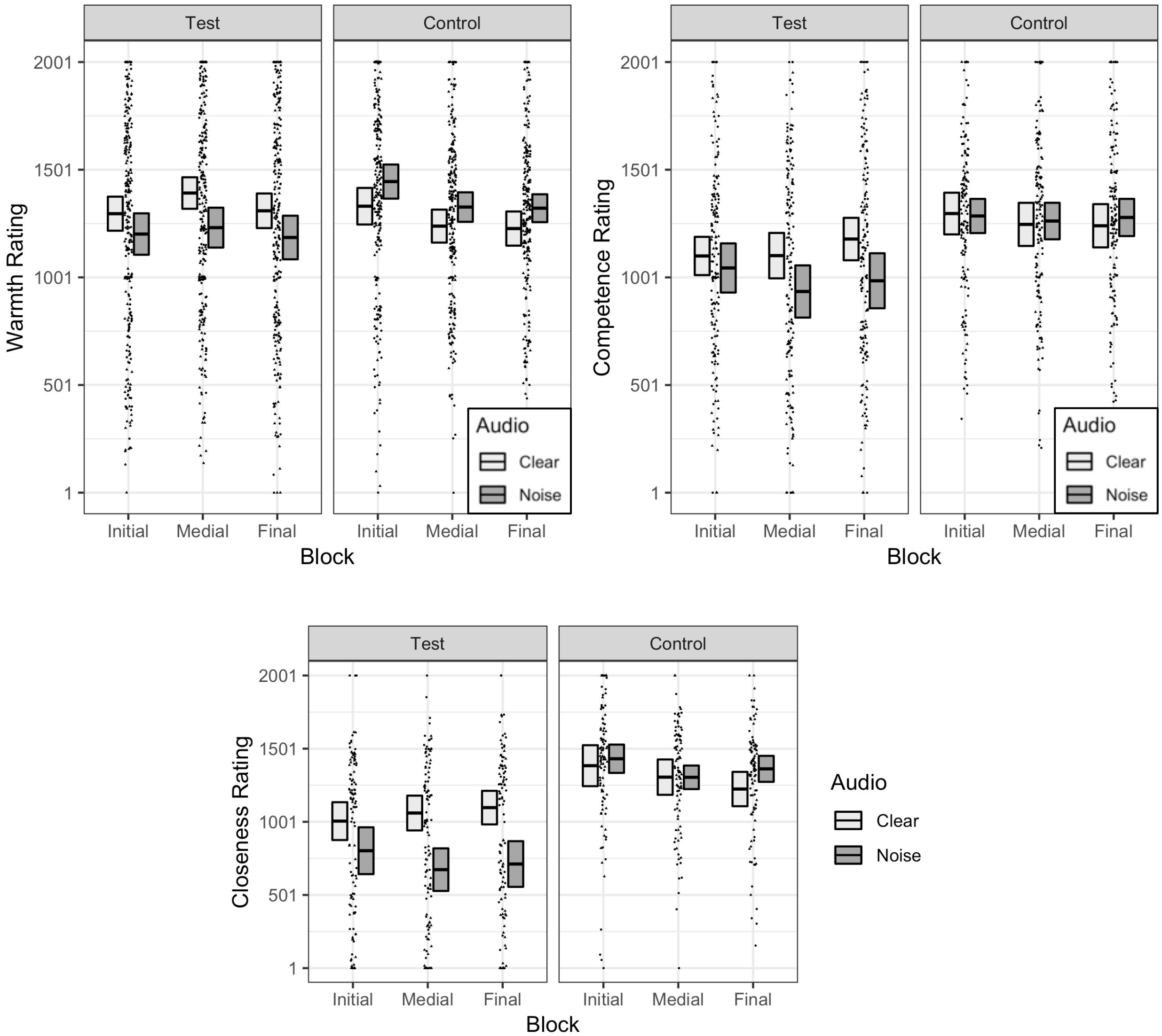

Figure 3.8 Warmth (top left), Competence (top right), and Closeness (bottom) ratings by Block, Audio, and Condition. The effect of Condition was significant for Warmth and Closeness and marginally significant for Competence. The interaction of Condition and Audio was significant for Closeness and marginally significant for Warmth and Competence. The effect of Initial-Medial Block was significant for Closeness and marginally significant for Warmth and Competence. The interaction effect of Initial-Medial Block and Condition was significant for Warmth ratings.

Dependent Variable	Fixed Effect (Predictor Variable)	β	t	p
Warmth Rating	**(Intercept)**	**1034.177**	**21.052**	**< 0.001**
	Block 1 Initial to Medial	*-40.956*	*-1.722*	*0.085*
	Block 2 Medial to Final	26.947	0.947	0.344
	Test Condition	**-260.405**	**-3.474**	**0.001**
	Audio Noise	-9.307	-0.156	0.876
	Baseline Rating	**0.214**	**9.367**	**< 0.001**
	Block 1 * Condition	**173.962**	**3.912**	**< 0.001**
	Block 2 * Condition	-34.748	-0.785	0.433
	*Condition * Audio*	*-227.542*	*-1.908*	*0.061*
	Condition * Baseline Rating	**0.157**	**4.201**	**< 0.001**
Competence Rating	**(Intercept)**	**822.416**	**14.527**	**< 0.001**
	Block 1 Initial to Medial	*-58.733*	*-1.845*	*0.065*
	Block 2 Medial to Final	35.431	0.940	0.348
	Test Condition	*-171.869*	*-1.869*	*0.062*
	Audio Noise	-60.498	-1.091	0.279
	Baseline Rating	**0.278**	**8.938**	**< 0.001**
	*Condition * Audio*	*-206.293*	*-1.860*	*0.067*
	Condition * Baseline Rating	-0.010	-0.174	0.862
Closeness Rating	**(Intercept)**	**1113.777**	**23.107**	**< 0.001**
	Block 1 Initial to Medial	**-70.874**	**-2.104**	**0.036**
	Block 2 Medial to Final	22.554	0.572	0.567
	Test Condition	**-445.042**	**-5.498**	**< 0.001**
	Audio Noise	-127.213	-1.572	0.121
	Condition * Audio	**-378.339**	**-2.337**	**0.022**

Table 3.10 Model statistics and parameters for Warmth, Competence, and Closeness comparing Initial, Medial, and Final blocks. Bold effects are significant at $p < 0.05$. Italic effects are marginally significant at $p < 0.1$.

3.9.5 The Relationship Between Fluency and Attitudes

Warmth, Competence, and Closeness rating were predicted to correlate negatively with objective fluency measured by Edit Distance, and positively with subjective fluency measured by Intelligibility rating. The results of the model construction and selection procedure testing the effects of Edit Distance and Intelligibility rating on attitudes (summing across data from the Initial, Medial, and Final blocks) did not uphold the prediction for Edit Distance. However, as predicted, Intelligibility rating (Figure 3.9) had a significant positive effect on Warmth rating ($\beta = 0.248$, $t = 4.372$, $p < 0.001$; top-left panels), Competence rating ($\beta = 0.279$, $t = 4.757$, $p < 0.001$; top-right panels) and Closeness rating ($\beta = 0.453$, $t = 6.319$, $p < 0.001$; bottom panels). For Closeness rating, there was also a significant interaction of Intelligibility with Condition ($\beta = 0.307$, $t = 2.144$, $p = 0.033$), indicating that the relationship between Closeness and Intelligibility rating was more positive for the non-native guise than the native guise. This effect is driven by one participant in the Control × Clear condition with outlying low Intelligibility rating. When that participant's data are excluded, the interaction of Intelligibility and Condition is no longer significant. Parameter estimates and statistics for these models are shown in Table 3.11.

Dependent Variable	Fixed Effect (Predictor Variable)	β	t	p
Warmth Rating	**(Intercept)**	**1060.919**	**17.392**	**< 0.001**
	Intelligibility	**0.248**	**4.372**	**< 0.001**
	Test Condition	**179.306**	**2.290**	**0.024**
	Audio Noise	75.173	1.226	0.224
	Condition * Audio	**-231.893**	**-2.028**	**0.047**
Competence Rating	**(Intercept)**	**906.978**	**14.700**	**< 0.001**
	Intelligibility	**0.279**	**4.757**	**< 0.001**
	Test Condition	49.973	0.656	0.513
	Audio Noise	32.229	0.563	0.575
	Condition * Audio	-142.072	-1.353	0.181
Closeness Rating	**(Intercept)**	**757.362**	**9.822**	**< 0.001**
	Intelligibility	**0.453**	**6.319**	**< 0.001**
	Test Condition	**-304.747**	**-1.976**	**0.049**
	Audio Noise	43.891	0.587	0.559
	Intelligibility * Condition	**0.307**	**2.144**	**0.033**
	*Condition * Audio*	*-263.544*	*-1.763*	*0.082*

Table 3.11 Model parameters and statistics for regressions of Warmth, Competence, and Closeness rating over Intelligibility rating. Bold effects are significant at $p < 0.05$. Italic effects are marginally significant at $p < 0.1$.

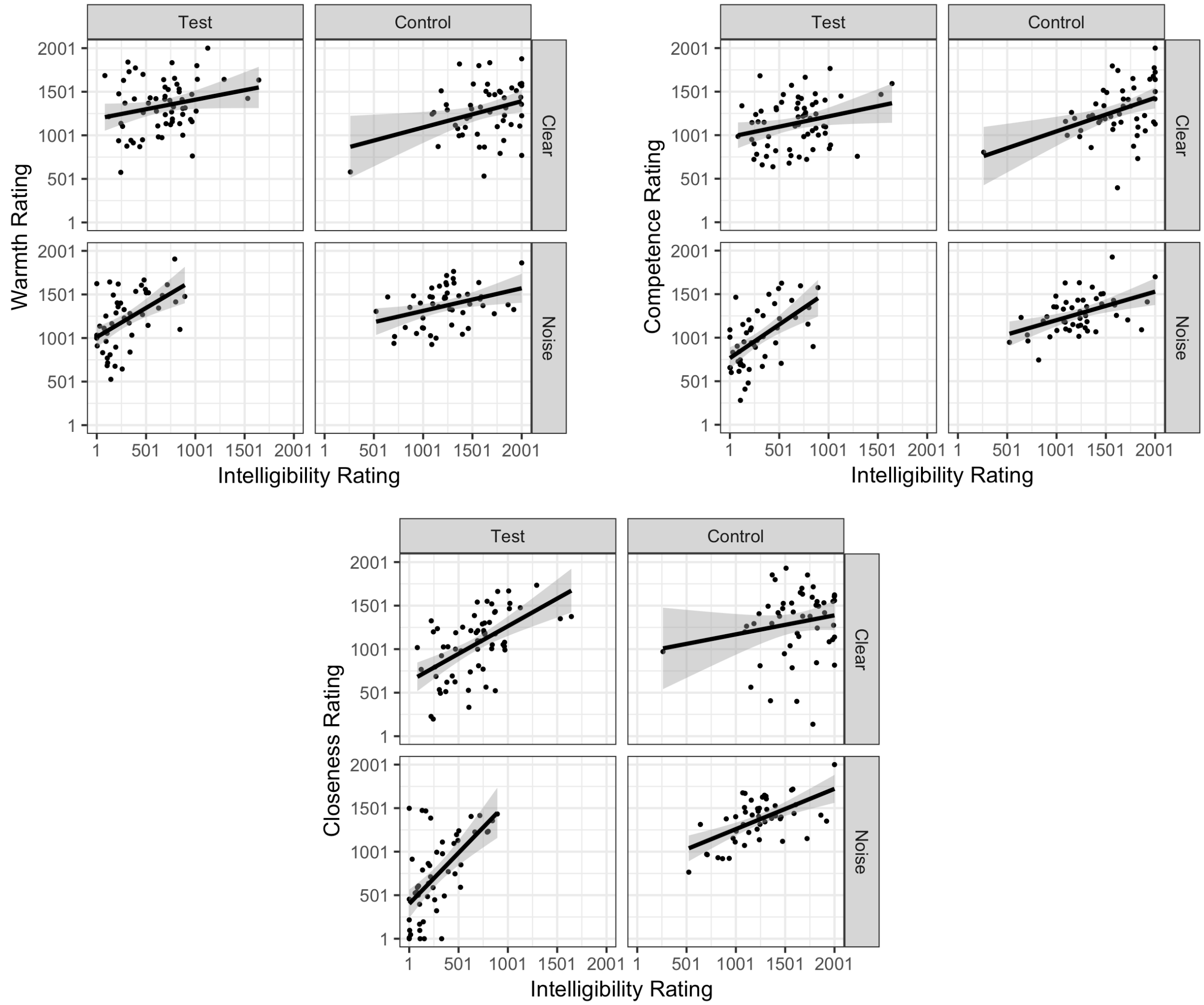

Figure 3.9 Warmth (top left), Competence (top right), and Closeness (bottom) rating by Intelligibility rating summing across the Initial, Medial, and Final blocks. The effect of Intelligibility rating was significant for each dependent variable.

3.9.6 The Relationship Between Changes in Fluency and Attitude

The results of Experiment 2 presented thus far, unlike the results of Experiment 1, do not provide evidence that either objective or subjective processing fluency improves over time as listeners gain experience with a non-native accent (section 3.9.3). Attitudes toward the speaker in the non-native guise also did not improve over time (section 3.9.4). However, the analysis in section 3.9.5 of the relationship between processing fluency measures and attitude results summed across blocks indicates that intelligibility, though not transcription accuracy, is positively correlated with evaluations of warmth, competence, and closeness.

Increased fluency over time is also predicted to correlate with improvement in speaker evaluation. Contrary to predictions, there was no evidence for a negative correlation between change in Edit Distance and change in Warmth, Competence, or Closeness rating, whether from the Initial to Medial or from the Medial to Final block. Although contrary to predictions, this result is consistent with the lack of evidence for a relationship between Edit Distance and attitude rating summing across blocks.

However, models of change in Warmth, Competence, and Closeness rating regressed over change in Intelligibility rating partially support the prediction that increased Intelligibility correlates positively with improvement in attitude. From the Initial to Medial block (Figure 3.10), increased Intelligibility rating had a significant positive correlation with increased Competence ($\beta = 0.258$, $t = 2.603$, $p = 0.011$) and Closeness rating ($\beta = 0.330$, $t = 3.314$, $p = 0.003$). For Closeness, the effect of Intelligibility was modulated by a significant three-way interaction with Condition and Audio due to the absence of a positive correlation in the Control × Clear condition (see Table 3.12 and Figure 3.10, upper right-most panel). No significant effects of increased Intelligibility rating on increased Warmth rating were found from the Initial to Medial block. Parameter estimates and statistics for the models of fluency and attitude change from the Initial to Medial block are shown in Table 3.12.

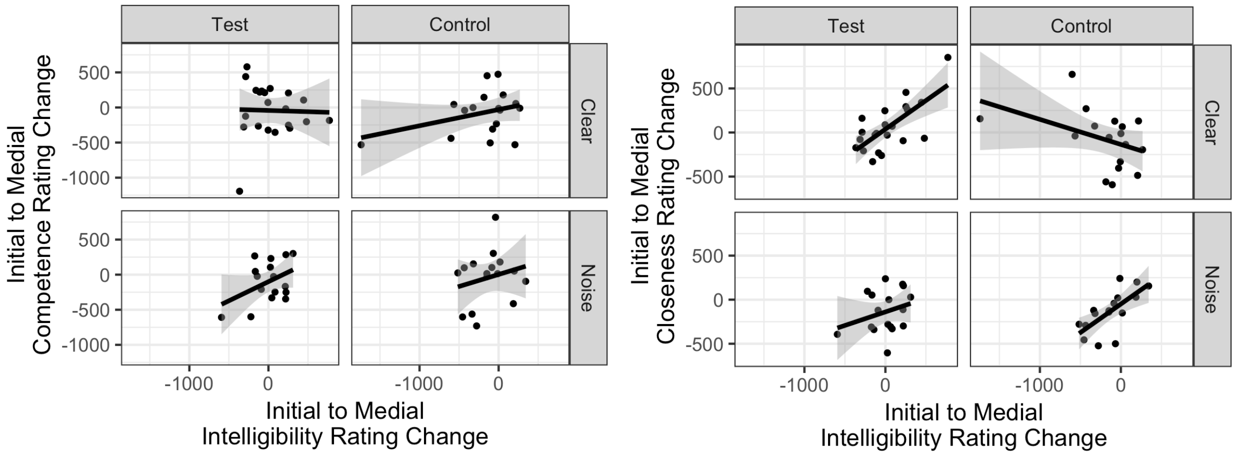

Figure 3.10 Initial to Medial Competence (left) and Closeness (right) rating change by Intelligibility rating change.

Dependent Variable	Fixed Effect (Predictor Variable)	β	t	p
Initial to Medial Competence Rating Change	(Intercept)	-33.373	-1.028	0.308
	Intelligibility Rating Change	**0.258**	**2.603**	**0.011**
	Test Condition	-29.904	-0.454	0.651
	Audio Noise	-32.038	-0.506	0.614
	Condition * Audio	-162.423	-1.280	0.205
Initial to Medial Closeness Rating Change	**(Intercept)**	**-71.544**	**-2.384**	**0.020**
	Intelligibility Rating Change	**0.330**	**3.134**	**0.003**
	Test Condition	41.840	0.697	0.488
	Audio Noise	-44.756	-0.746	0.459
	Intelligibility * Condition	0.297	1.409	0.164
	Intelligibility * Audio	0.295	1.402	0.166
	Condition * Audio	**-263.366**	**-2.194**	**0.032**
	Intelligibility * Condition * Audio	**-1.275**	**-3.026**	**0.004**

Table 3.12 Model parameters and statistics for Initial-Medial attitude change and fluency change. Bold effects are significant at $p < 0.05$. Italic effects are marginally significant at $p < 0.1$.

Modeled results comparing changes from the Medial to Final block (Figure 3.11) also partially support the prediction that increased Intelligibility correlates positively with improved attitude. Increased Intelligibility rating had a significant positive correlation with increased Warmth ($\beta = 0.187$, $t = 2.102$, $p = 0.039$) and Closeness ($\beta = 0.320$, $t = 2.683$, $p = 0.009$) rating.

The main effect of change in Intelligibility rating in the model of change in Competence rating was not significant. Parameter estimates and statistics for the models of fluency and attitude change from the Initial to Medial block are shown in Table 3.13.

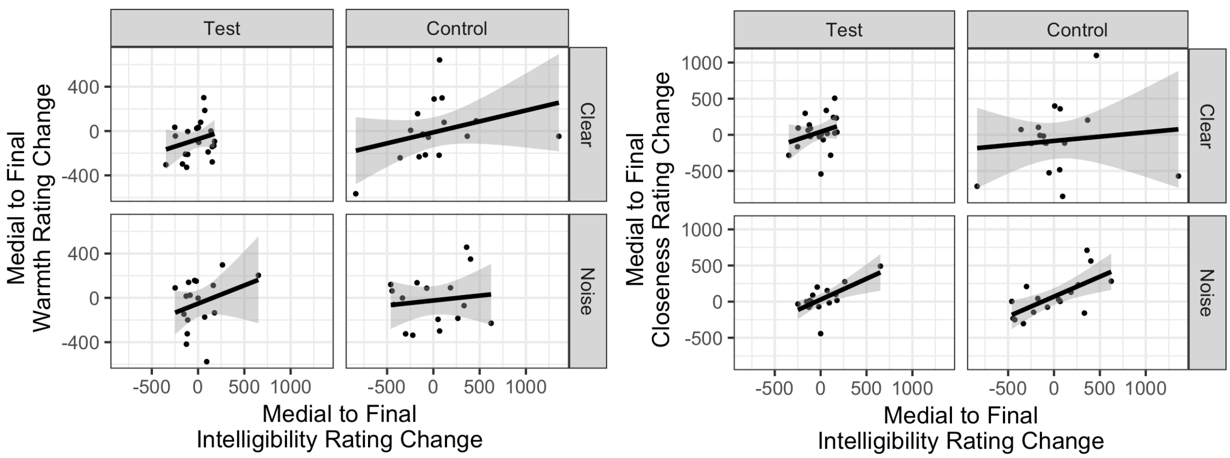

Figure 3.11 Medial to Final Warmth (left) and Closeness (right) rating change by Intelligibility rating change.

Dependent Variable	Fixed Effect (Predictor Variable)	β	t	p
Medial to Final Warmth Rating Change	(Intercept)	-39.662	-1.506	0.137
	Intelligibility Rating Change	**0.187**	**2.102**	**0.039**
	Test Condition	-44.637	-0.847	0.400
	Audio Noise	6.841	0.130	0.897
	Condition * Audio	36.402	0.345	0.731
Medial to Final Closeness Rating Change	(Intercept)	13.166	0.372	0.711
	Intelligibility Rating Change	**0.320**	**2.683**	**0.009**
	Test Condition	47.230	0.668	0.507
	Audio Noise	74.801	1.058	0.294
	Condition * Audio	-164.109	-1.158	0.251

Table 3.13 Model parameters and statistics for Medial-Final attitude change and fluency change. Bold effects are significant at $p < 0.05$.

While many of the effects of Intelligibility change on attitude change from the Initial to Medial and Medial to Final blocks are significant, the magnitude of the effects is not large.

Additionally, Figure 3.10 and Figure 3.11 show that, for several participants in each condition, intelligibility rating increased while attitude rating decreased or vice versa. The data clearly indicate that change in intelligibility and attitude were positively associated, but with considerable variation across participants.

3.9.7 The Relationship of Subjective Fluency with Objective Fluency and Attitudes

The results presented in sections 3.9.5 and 3.9.6 do not provide evidence that objective fluency (Edit Distance) and attitudes are related but do provide evidence that subjective fluency (Intelligibility rating) and attitudes are related. These patterns match those in Experiment 1 and suggest that Intelligibility rating may reflect or be strongly mediated by social judgment. To test the relationship of subjective fluency with objective fluency versus attitude rating, models were constructed as described in section 3.8.6 for each of the three Intelligibility items (see section 3.2). Edit Distance and Closeness rating both correlated significantly with Intelligibility ratings, while Warmth and Competence rating did not and are not included in the final models.

The relationships of Edit Distance and Closeness rating with Transcribe, Speaks, and Comprehend ratings are shown in Figure 3.12, summing across data from the Initial, Medial, and Final blocks. Consistent with predictions, there were significant negative effects of Edit Distance on Transcribe item ratings ($\beta = -20.092$, $t = -2.393$, $p = 0.018$; Figure 3.12, top-left panels) and Comprehend item ratings ($\beta = -15.158$, $t = -1.985$, $p = 0.048$; Figure 3.12, bottom-left panels). However, the *p*-values for these effects were large and, as in Experiment 1, there were highly significant correlations between Closeness ratings and each Intelligibility item rating (see Table 3.14). While these results provide tentative evidence that the subjective experience of processing fluency is related to objective processing fluency, there is stronger evidence that subjective processing fluency is related to social perception.

The three-way interactions of Closeness rating with Condition and Audio indicate that the relationship between Closeness rating and Intelligibility item ratings is more positive in Noise than Clear Audio, but this difference is larger in the Control condition than the Test condition. This may indicate that perceptions of intelligibility are more sensitive to judgments of social closeness, or vice versa, when intelligibility of a native-accented speaker is hindered by noise, but less so when intelligibility is hindered by a non-native accent. These patterns will be discussed in Chapter 4. A summary of the parameter estimates and statistics for these models is shown in Table 3.14.

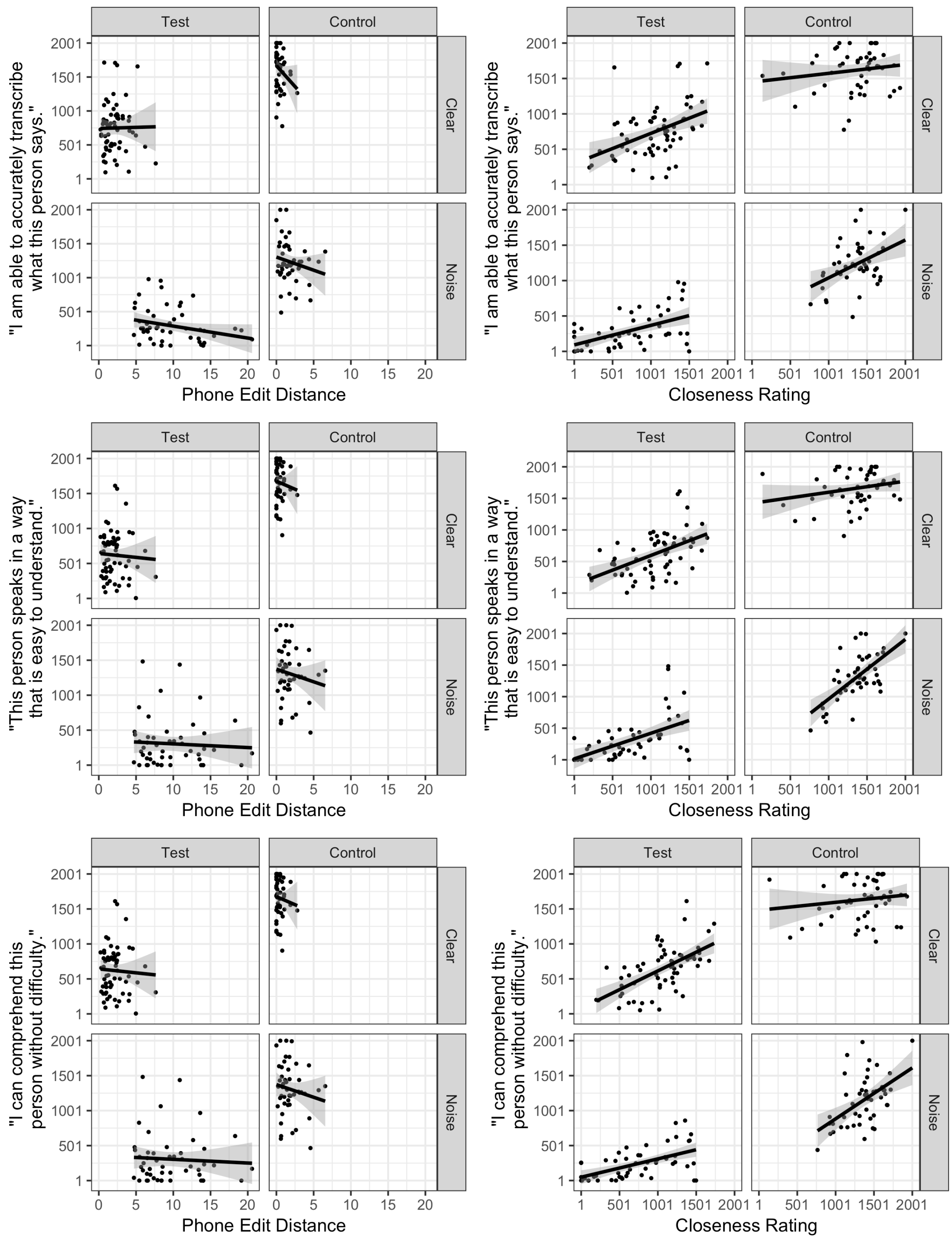

Figure 3.12 Transcribe (top), Speaks (middle), and Comprehend (bottom) ratings by Edit Distance (left) and Closeness rating (right). The effect of Edit Eistance was significant for Transcribe and Comprehend ratings. The effect of Closeness rating was significant for all three Intelligibility items. The three-way interaction of Closeness, Condition, and Audio was significant for Speaks and Comprehend ratings and marginally significant for Transcribe ratings.

Dependent Variable	Fixed Effect (Predictor Variable)	β	t	p
Transcribe Rating	**(Intercept)**	**651.805**	**7.284**	**< 0.001**
	Edit Distance	**-20.092**	**-2.393**	**0.018**
	Closeness rating	**0.340**	**5.319**	**< 0.001**
	Test Condition	**-599.590**	**-3.537**	**0.001**
	Audio Noise	**-554.554**	**-3.324**	**0.001**
	Closeness * Condition	-0.047	-0.368	0.713
	Closeness * Audio	0.191	1.503	0.134
	Condition * Audio	**802.252**	**2.437**	**0.016**
	*Closeness * Condition * Audio*	*-0.492*	*-1.930*	*0.055*
Speaks Rating	**(Intercept)**	**432.937**	**5.043**	**< 0.001**
	Edit Distance	-12.362	-1.557	0.121
	Closeness rating	**0.502**	**8.148**	**< 0.001**
	Test Condition	**-548.349**	**-3.363**	**0.001**
	Audio Noise	**-623.962**	**-3.898**	**< 0.001**
	Closeness * Condition	-0.158	-1.292	0.198
	Closeness * Audio	**0.290**	**2.366**	**0.019**
	Condition * Audio	**1244.463**	**3.936**	**< 0.001**
	Closeness * Condition * Audio	**-0.751**	**-3.062**	**0.003**
Comprehend Rating	**(Intercept)**	**470.357**	**5.775**	**< 0.001**
	Edit Distance	**-15.158**	**-1.985**	**0.049**
	Closeness rating	**0.425**	**7.290**	**< 0.001**
	Test Condition	**-597.662**	**-3.870**	**< 0.001**
	Audio Noise	**-624.477**	**-4.111**	**< 0.001**
	Closeness * Condition	-0.090	-0.776	0.439
	Closeness * Audio	0.182	1.569	0.118
	Condition * Audio	**1357.466**	**4.529**	**< 0.001**
	Closeness * Condition * Audio	**-0.860**	**-3.707**	**< 0.001**

Table 3.14 Model parameters and statistics for Intelligibility item regressions. Bold rows indicate significance at $p < 0.05$. Italic effects are marginally significant at $p < 0.1$.

3.10 Discussion

Chapter 4 will provide a full summary and discussion of hypotheses, predictions, and results from both experiments. This section provides a more focused assessment of selected patterns in the results of Experiment 2 and their relation to Experiment 1 results. Contrary to predictions but consistent with Experiment 1 results, the results from Experiment 2 do not provide evidence to support the hypothesis that objective processing fluency, as measured by transcription accuracy, improves with experience with a non-native accent. Rather, transcription accuracy improved for both the native and non-native guises. Also contrary to expectations, subjective processing fluency, as measured by intelligibility ratings, was not shown to improve with non-native accent experience. Rather, again consistent with the results of Experiment 1, intelligibility ratings decreased over time in the control condition but not the test condition. In the discussion of Experiment 1 results (section 2.12), it was speculated that the decrease in subjective processing fluency in the control condition may be due to fatigue resulting from the experimental task that was counteracted by increased objective fluency in the test condition, as measured by dilation and transcription time in Experiment 1. This interpretation is consistent with the earlier observed decrease in Experiment 2 (from the initial to medial block) than Experiment 1 (from the medial to final block). Experiment 2, unlike Experiment 1, began with favorability ratings for 39 adjectives and Baseline speaker ratings for 39 adjectives (versus 13 in Experiment 1). By the time participants reached the medial block intelligibility rating, the additional time and tasks in Experiment 2 may have resulted in fatigue that was sufficient to negatively influence intelligibility ratings, whereas in Experiment 1, the effects of fatigue became evident only in the final block. However, there is no evidence for a greater increase in objective fluency in the test condition than the control condition in Experiment 2 (as measured by edit distance). Objective fluency may still have improved without measurably affecting transcription accuracy, as was observed based on dilation, transcription time, and edit distance measures in Experiment 1, but this interpretation remains highly speculative.

Experiment 2, like Experiment 1, provides evidence that, consistent with predictions, high intelligibility rating is associated with positive attitude rating, and that when Intelligibility rating improves over time attitude rating also tends to improve (though this effect varied across participants). These patterns were found for Warmth, Competence, Closeness evaluations. However, as in Experiment 1, there was no evidence that attitude rating was linked to objective fluency: ratings were not more positive when transcription was more accurate nor did they improve over time as transcription errors decreased.

Finally, the addition of speech-shaped noise to stimuli for half of the participants in Experiment 2 provides a second manipulation of processing fluency that is useful for testing the hypothesis that fluency and attitudes are related. Overall, the results from Experiment 2 indicate that noise-related disfluency by itself did not have an effect on speaker evaluation, but that noise interacted with the accent in its effects on processing fluency and speaker evaluation. In clear audio, warmth ratings were higher for the target speaker in both the non-native and native guise than ratings for the baseline speaker. However, in noise, warmth ratings were higher for the target speaker than the baseline speaker only in the native guise. For the target speaker in the initial, medial, and final blocks, warmth ratings were higher for the native guise than the non-native guise in noise but there was no difference in ratings for the two guises in clear audio. For competence and closeness ratings, the native guise was rated comparably in noise and clear audio, but the non-native guise received lower ratings in noise than clear audio. The interactions suggest that disfluency caused by noise was attributed to the non-native accent but not the native accent, resulting in more negative attitudes toward the non-native accent. The effects of noise will be further discussed in Chapter 4, in connection with the questions of whether participants perceive fluency differently for native and non-native accents, and how the perception of fluency is related to social judgment.

CHAPTER 4

General Discussion, Conclusions, and Implications

The two experiments reported in Chapters 2 and 3 investigated the role of processing fluency in shaping attitudes toward non-native speakers as well as adaptation to an accent over time and its effects on both fluency and attitudes. The results contribute to both perceptual adaptation and language attitudes research by measuring physiological, behavioral, and self-reported indicators of fluency in a listening and transcription task and relating those measures to attitude changes in real time as listeners gain experience with an unfamiliar accent.

The main hypotheses and related findings of this dissertation are summarized in Table 4.1 and Table 4.2. Overall, evidence from Experiment 1 indicates that objective speech processing fluency, as measured by decreased pupil dilation and transcription time, improves over time as listeners adapt. The positive correlation of pupil dilation with transcription errors in Experiment 1 indicates that transcription accuracy is related to listener effort, but transcription accuracy unexpectedly did not improve over time to a greater degree for non-native accented speech than native-accented speech in either experiment. The evidence for a relationship between objective and subjective processing fluency (the latter as measured by intelligibility rating) is tentative and restricted to Experiment 2 results, and there is strong evidence indicating that subjective processing fluency may be more closely related to perceptions of social closeness (Table 4.1). Neither experiment provided evidence that subjective processing fluency and attitudes improve over time in relation to the improvement in objective processing fluency (Table 4.2). These findings suggest that the relationship between the ease with which a listener can comprehend speech and their attitudes toward the speaker is not direct, but may be mediated by the listeners' subjective perception of fluency and naïve theories about the source and meaning of ease.

Experiment 1	Experiment 2
• Pupil dilation and transcription time (but not transcription errors) decreased over trials, more for test than control participants. • Pupil dilation correlated positively with transcription errors and transcription time on individual trials.	• Transcription errors decreased over trials, but the decrease did not differ between test and control conditions nor between noise and clear audio
• Intelligibility ratings decreased from medial to final block for control but not test participants.	• Intelligibility ratings decreased from initial to medial block for control but not test participants (in both noise and clear audio).
• No evidence that objective and subjective fluency measures are related.	• Subjective fluency ratings of transcription accuracy and comprehension ease increased with objective transcription accuracy, though the effect was not strongly significant.
• Subjective fluency correlated positively with closeness.	• Subjective fluency was strongly positively correlated with closeness ratings. The relation between closeness and subjective fluency was more positive in noise than clear audio in the control condition.

Table 4.1 Summary of results related to the hypothesis that processing fluency will increase over time as listeners gain experience with a non-native accent.

	• Competence ratings marginally decreased from initial to medial block for control participants but not test participants	• Warmth ratings decreased from initial to medial block for control participants but not test participants (in both noise and clear audio)
Attitude change over time	[No noise condition]	• For native-accented speech, no evidence that attitude ratings were negatively affected by noise. For non-native accented speech, warmth, competence, and closeness ratings were negatively affected by noise. • No evidence that attitudes improved over time for native or non-native accented speech in noise.
Relationship between attitudes and processing fluency	• Intelligibility correlated positively with warmth, competence and closeness. • No evidence that objective fluency and attitudes are related	• Intelligibility correlated positively with warmth, competence, and closeness. • No evidence that objective fluency and attitudes are related
	• Intelligibility change from medial to final block correlated positively with closeness change and marginally positively with competence change	• Intelligibility change from initial to medial block correlated positively with competence and closeness change. • Intelligibility change from medial to final block correlated positively with warmth and closeness change.

Table 4.2 Summary of results related to the hypothesis that attitudes toward a non-native speaker (and toward a native speaker in noise) will improve over time, in relation to increased fluency.

Before turning to these findings and their implications for the main hypotheses, this chapter first briefly discusses an unexpected result regarding perceptions of warmth in non-native speakers relative to native speakers: while fluency measures and competence and closeness ratings were lower overall for the non-native than the native guise in both experiments, there was no evidence of a difference in warmth ratings. Comparison of these results with previous research on processing fluency and language attitudes raises theoretical and methodological considerations relevant to future research on language attitudes and attitude change (section 4.1). After this, the discussion returns to the main hypotheses concerning improvement of fluency over time and the theorized link between processing fluency and language attitudes, examining in particular the lack of evidence that objective measures of fluency are related to participants' reported fluency and attitudes toward the speaker (section 4.2). These results, considered in connection with other processing fluency and attitudes research, motivate a revised model of the relationship between processing fluency and language attitudes (section 4.3). The chapter then turns to avenues for future research, including methodological considerations, suggested by the present findings (section 4.4), and closes with practical conclusions for speakers and listeners as they encounter unfamiliar accents (section 4.5).

4.1 Attitudes Toward Non-Native Speakers

Differences in attitudes toward the non-native and native accented guises in this study were not fully consistent with previous research examining attitudes and processing fluency. This section addresses those inconsistencies and discusses the measurement and interpretation of attitudes toward non-native speakers. The results for attitude change over time are discussed in section 4.2.3.

Contrary to predictions, neither experiment provided evidence that evaluations of warmth are more negative for a speaker with a non-native accent than a native accent. Evaluations of competence and closeness, however, were more negative for a non-native than native-accented speaker. Comparison of this pattern of results with previous fluency and attitudes research suggests the need for caution in describing non-native speakers as a unified group, given the other identities and stereotypes that often intersect with linguistic identity. Comparison also raises methodological considerations for the use of adjective ratings to measure language attitudes in future research.

The warmth and competence rating patterns here are virtually the opposite of those reported by Dragojevich and Giles (2016): in that study, participants who heard a female speaker in a Punjabi English guise rated her lower on solidarity (warmth) than participants who heard an American English guise, but ratings on status (competence) did not differ between groups. However, the patterns found here may be consistent with the results reported by Dragojevic, Giles, Beck, and Tatum (2017). In that experiment, the researchers gathered participant ratings for mild- and heavy-accented guises of a male Punjabi English speaker and a male native-Chinese English speaker. For both speakers, status (competence) ratings, but not solidarity (warmth), were lower for the heavy accent than the mild accent. While that study compared two guises that differed in degree of accent, the effect of degree of accent on competence but not warmth ratings matches the effects of non-native accent in the present research, which also used a male speaker. The inversion of warmth and competence differences may be suggestive of an influence of the sex or gender of the speaker in attitudes toward accented speakers, although this interpretation is highly speculative. Future research on attitudes toward non-native accents, though, may benefit from accounting for speaker gender.

The different effects of accent on perceptions of warmth and closeness, in relation to previous research, may also be related to non-gender-based accent stereotypes. Dragojevic and colleagues used fairly recognizable accents and in most cases told participants what accent they would hear. In contrast, the present study used an accent chosen because listeners found it difficult to identify or categorize. While this was intended to reduce the effect of accent stereotypes on evaluation, it is possible that listeners nonetheless inferred a vague stereotype or category for the non-native guise. Responses to debriefing questions indicated that participants were clearly interested in identifying why the speaker spoke the way he did, most commonly where he was from and how he learned English.[8] However, only seven out of the 61 participants in the test condition whose data were included in the analysis made any conclusions about the speaker's origin (those guesses included Asia, north Africa, Germanophone Europe, and the Caribbean). The stereotypes that most participants likely inferred from the speaker's accent were "non-native" and "non-local," and what effect these stereotypes may have had on attitude is unclear. Additional research is needed to identify how more- or less-specific stereotypes related to native language or place of origin influence evaluations of the speaker.

Another methodological difference between the present research and the experiments conducted by Dragojevich and Giles (2016) that may have influenced the pattern of results for warmth and competence ratings is the adjectives used to measure speaker evaluations, and these too may have resulted in different evaluation patterns. Dragojevich and Giles used exclusively favorable adjective descriptions to measure solidarity (friendly, nice, pleasant, honest, sociable) and status (intelligent, educated, smart, competent, successful). In the present study, 15 out of 24 warmth adjectives and six out of 15 competence items were unfavorable for a majority of participants in Experiment 2 (see Table 3.3), and there is some evidence in previous attitude research that suggests that unfavorable trait evaluation may differ systematically from favorable trait evaluation. For example, Bergsieker et al. (2012) presented participants with descriptions of

[8] Five participants speculated that the speaker might have a speech disorder. While this might suggest that the recorded accent was perceived as unnatural or inauthentic, none of the participants (including those who provided pilot data and feedback on the experimental design) expressed any suspicion that the speaker was imitating or acting the accent.

an individual's warmth and competence behavior in which one trait was described positively while the other was described negatively. In follow-up evaluations, participants tended to avoid descriptions of the individual that used unfavorable traits more than descriptions that used favorable traits, even when an unfavorable description was accurate given the described behavior.

If participants in the current study shared this bias, ratings on unfavorable adjectives would tend toward "Not at all" ratings, while ratings on favorable adjectives would not show an opposing tendency toward "Very" ratings, leading to generally more positive evaluations measured by unfavorable versus favorable adjectives. A post-hoc model testing differences in ratings by adjective favorability in Experiment 2 (Figure 4.1) indicated that evaluations are indeed more positive when measured by unfavorable warmth and competence adjectives. For both warmth and competence ratings, this effect interacted with Condition such that the difference between unfavorable and favorable adjective ratings was larger for the non-native guise (Test) than the native guise (Control). (Model estimates and statistics are shown in Table 4.3.) Including Adjective Valence and its interaction with Condition as effects in models of Warmth and Competence ratings in Experiment 2 did not change the patterns of results reported in Chapter 3, so it is unlikely that the present results would be different if only positive adjectives were included. Whether the effect of adjective valence could account for the opposing patterns in this study and Dragojevic and Giles' (2016) study is unclear, but it appears that the adjectives by which listeners evaluate non-native speakers should be taken into account when interpreting speaker evaluations.

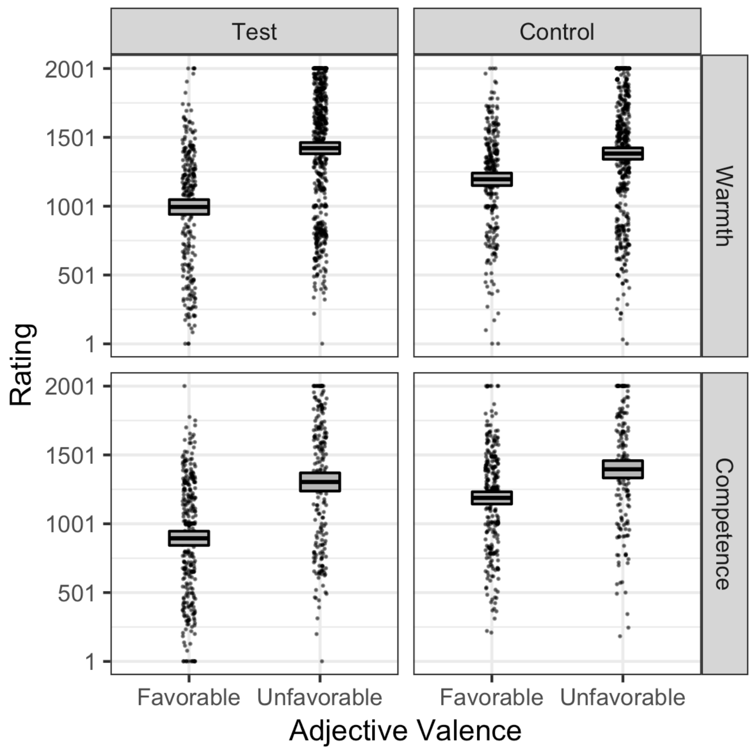

Figure 4.1 Ratings for warmth and competence traits by adjective valence in the initial, medial, and final blocks of Experiment 2. Crossbars indicate mean and 95% confidence intervals. The effect of Valence and the interaction of Valence and Condition were significant for both warmth and competence ratings.

Dependent Variable	Fixed Effect (Predictor Variable)	β	t	p
Warmth Rating	**(Intercept)**	**1255.544**	**42.666**	**< 0.001**
	Unfavorable Valence	**289.817**	**14.155**	**< 0.001**
	Test Condition	-84.466	-1.435	0.156
	Valence * Condition	**232.415**	**5.676**	**< 0.001**
Competence Rating	**(Intercept)**	**1201.215**	**42.675**	**< 0.001**
	Unfavorable Valence	**294.262**	**11.897**	**< 0.001**
	Test Condition	**-190.220**	**-3.379**	**0.001**
	Valence * Condition	**195.274**	**3.948**	**< 0.001**

Table 4.3 Model estimates and statistics comparing ratings by adjective valence in the initial, medial, and final blocks of Experiment 2. Bold effects are significant at $p < 0.05$.

4.2 Main Hypotheses and Research Questions

4.2.1 Increased Ease and Perceptions of Ease with Perceptual Adaptation to Non-Native Speech

Experiment 1 showed clear evidence that processing becomes less effortful—that is, some measures of objective fluency improve—as listeners gain experience with a non-native accent. Multiple studies have shown evidence that listeners become more accurate in their comprehension of a non-native speaker's intended utterance over time (Baese-Berk et al., 2013; Bradlow & Bent, 2008; Clarke & Garrett, 2004; Sidaras et al., 2009), but this research has not examined whether this improvement reflects easier processing or increased effort (i.e., recruiting additional cognitive resources to accurately comprehend unfamiliar speech patterns). Experiment 1 added to this research by measuring pupil dilation, a physiological response associated with cognitive effort, during sentence comprehension. The observed decrease in pupil dilation and transcription time over trials (Figure 2.4) is consistent with the interpretation that, as listeners perceptually adapt to an accent with increased experience with that accent, the cognitive effort involved in comprehension decreases.

However, findings for transcription accuracy were not as predicted. There was no evidence that transcription accuracy improved over time in Experiment 1 (Figure 2.5), though this may have been due to near-ceiling accuracy in both the test and control conditions throughout that experiment. While transcription accuracy did improve over time in Experiment 2, there were no differences in accuracy improvement between the non-native (test condition) and native (control condition) guises or between clear audio and audio mixed with noise (Figure 3.6). Experiment 1 did show a clear positive relationship between pupil dilation and transcription errors for individual sentence transcription trials (Figure 2.10), supporting the conclusion that comprehension errors are a useful indicator of cognitive effort in speech processing. Consistent with this relationship, transcription accuracy was worse for the non-native guise than the native guise in both experiments (Figures 2.5 and 3.6), and in noise than clear audio in Experiment 2 (Figure 3.6). In light of these patterns, the absence of evidence for an increase in accuracy with adaptation to the non-native accent may indicate that the dependent variable (phone edit distance) and statistical analysis were insufficient to detect a change in accuracy over time. Phone edit distance is a novel measure of accuracy in adaptation research, designed to detect miscomprehension at the phonetic level using automated scoring, and previous studies have primarily used the proportion of words in a sentence that were transcribed or repeated correctly as the dependent variable. Additional research is needed to determine more precisely the factors that influence the relation between processing fluency and comprehension accuracy in perceptual adaptation to accented speech.

While pupil dilation and transcription time (in Experiment 1) clearly indicate increased objective ease over time, neither experiment showed direct evidence for increased subjective ease. In Experiment 1, subjective fluency decreased from the middle to the end of the experiment in the control group but not the test group (Figure 2.6). Experiment 2 results showed the same pattern but with the decrease occurring from the beginning to the middle of the experiment (Figure 3.7). These unexpected patterns may be interpreted in several ways. One interpretation that may be tentatively discounted is that ratings near the upper limit of the intelligibility rating scale in the control condition allowed ratings to decrease but not increase. In Experiment 2, intelligibility ratings for the control condition were near ceiling in clear audio, but not in noise. If the decrease from initial to medial block were due to a general drift in ratings away from from the upper bound of the scale, it would only be expected to occur in the control × clear and not in the control × noise condition. Contrary to this interpretation, the decrease from the Initial Block to the Medial block did not differ between Noise and Clear audio (see Figure 3.7). A related interpretation is that intelligibility ratings did not decrease in the test condition because they were already near the lower limit of the response scale, but this also seems unlikely. If intelligibility ratings in the test condition were limited by the lower bound of the scale, ratings in clear audio might be expected to decrease over time while ratings in noise would remain unchanged. However, there was no evidence for a change in ratings in the test condition in clear audio.

An alternate, and highly speculative, interpretation is that the decrease in subjective fluency measured by intelligibility ratings in the control condition was a general effect of the experimental setting or task (e.g., perhaps due to negative affect induced by fatigue or boredom). This interpretation is consistent with the earlier decrease in intelligibility ratings for the control group in Experiment 2, in which ratings decreased from the initial to medial block (Figure 3.7), than Experiment 1, in which ratings decreased from the medial to final block (Figure 2.6). Experiment 2 included substantially more adjective and speaker rating questions before the medial intelligibility rating than did Experiment 1. Assuming fatigue- or boredom-driven negative affect, the additional task demands early in Experiment 2 could have induced an earlier negative change in affect reflected in intelligibility ratings.

Negative affect is implicated in judgment and is associated with perceptions of disfluency, as discussed in Chapter 1 in the review of research linking disfluent processing to negative evaluation (see Forster et al., 2016; Winkielman et al., 2003). Therefore, it is plausible that a participant's general affect or mood would interact with fluency-driven affect in their perception of ease. If this is the case, the decrease over blocks for control participants but not test participants in both experiments may be evidence that test participants' fluency-driven affective response became more positive as objective fluency improved, counteracting the negative, task-driven affective response surmised from the decrease in subjective perceptions of fluency in the control condition. In Experiment 2, where the increase in transcription accuracy over time (see Figure 3.6) was not greater in the test condition than the control condition or in noise than clear audio, the interpretation based on supposed general task effects requires the assumption that objective fluency increased more for test versus control participants in Experiment 2, without an observable increase in transcription accuracy. This seems a plausible assumption, as it is consistent with objective fluency results in Experiment 1 (which showed decreased transcription time and pupil dilation over time), though there is no evidence to support it. (Recall that transcription time was not analyzed in Experiment 2 because short transcription times were ambiguously related to both high and low effort; see section 3.6.1.)

The interpretation that subjective fluency decreased more in the control condition than the test condition because objective fluency increased less in the control condition than the test condition also depends on a positive relationship between objective and subjective fluency, and the evidence for that relationship is limited. Objective fluency was positively related to (subjective) intelligibility ratings only in Experiment 2, in the correlation of transcription accuracy with *transcribe* and *comprehend* item ratings (Figure 3.12), and the statistical significance of this relationship was relatively weak ($p = 0.018$ for *transcribe* and $p = 0.049$ for *comprehend*; see Table 3.14), indicating that it may have been observed by chance. No evidence was found for a relationship of pupil dilation, transcription time, or transcription accuracy with intelligibility item ratings in Experiment 1.

In both experiments, however, there was a highly significant positive relationship between social closeness and ratings for all three intelligibility items (all p-values ≤ 0.006; see Table 2.11 and Table 3.14), and the change in intelligibility rating for each participant was positively related to change in closeness rating from the medial to final block in both experiments (Figure 2.9 and Figure 3.11) and from the initial to medial block in Experiment 2 (Figure 3.10). Closeness was not related to any of the objective measures of processing fluency in either experiment.

These patterns involving closeness suggest that participants' reported subjective perception of fluency was influenced more by social judgments and relatively dissociated from the experience of objective ease or effort comprehending the speaker, as measured by dilation, transcription time, and transcription accuracy. Under this interpretation, unchanged intelligibility ratings over time in the test condition are not connected to increased objective fluency, but reflect relatively stable evaluation of the non-native guise, perhaps due to a naïve theory held by listeners that influences their perception of fluency itself (see Winkielman et al., 2003). Such a naïve theory might hold that difficult comprehension is an attribute of the speaker rather than an attribute of the listening experience. By this account, objective disfluency on first hearing the non-native guise may have led test participants to form an impression of the speaker in which "difficult to understand" was a prominent trait—perhaps reinforced by the "difficult to understand" stereotype of non-native speakers (see Derwing et al., 2002; Lindemann, 2005)—and subsequent intelligibility reports reflected this impression of the speaker rather than ease or effort in processing.

In this interpretation, the decrease in intelligibility ratings over time in the control condition is the result of a change in participants' social evaluation of the speaker in the native guise, rather than a negative influence of the experimental task on affect. Social psychology research has shown that perceivers require more category-confirming information to confidently identify a person as a member of their own social group and little category-disconfirming information to identify them with an out-group (Leyens & Yzerbyt, 1992; Yzerbyt & Leyens, 1991). In the present research, as control (native guise) participants heard more sentences, the likelihood of encountering information perceived as inconsistent with the in-group increased. The observed decrease in intelligibility ratings would depend on a naïve theory, for example, that difficult comprehension is an out-group speaker attribute, which participants were prompted to infer upon encountering a stereotype-associated accent feature, or upon experiencing disfluency in comprehending a native-accented sentence in the medial or final block. This interpretation also assumes that disfluency caused by noise is not interpreted as an attribute of a native-accented speaker, as intelligibility rating patterns were the same in noise and clear audio (see Figure 3.7).

In summary, the pupil dilation and transcription time results from Experiment 1 provide clear evidence that objective fluency increases over time as listeners gain experience with a non-native accent. The intelligibility results from Experiment 1 and Experiment 2 do not clearly indicate that the increase in objective fluency was accompanied by an increase in subjective fluency and are consistent with two different and highly speculative interpretations. Under the assumption that the decrease in subjective fluency in the control condition was caused by negative affect (perhaps induced by fatigue or boredom from the experimental task), the unchanged intelligibility ratings in the test condition may indicate that increased objective fluency due to adaptation to the accent counteracted the negative influence of the task. However, there is no direct evidence that the experimental task had a generally negative influence on participants' processing experience. The second interpretation assumes a dissociation between objective fluency and participants' perception of fluency—based on stronger evidence that intelligibility is related to social closeness than to measures of objective fluency—and posits that changes in intelligibility ratings in the control condition reflected changes in social judgment over time. The dissociation of objective and subjective fluency has important implications for the hypothesis that processing fluency influences attitudes toward non-native speakers and will be returned to in the sections that follow.

4.2.2 Processing Fluency and Speaker Evaluation

Neither experiment provided evidence that objective fluency directly affected speaker evaluation. While objective fluency was lower (Figures 2.4, 2.5, and 3.6) and speaker evaluations were more negative on the competence and closeness dimensions (Figures 2.3, 2.7, 3.5, and 3.8) for the non-native guise than the native guise, there was no evidence that the two measures were related. Subjective fluency, on the other hand, was positively correlated with evaluations of warmth, competence and closeness in both experiments (Figures 2.8 and 3.9). The effects of noise in Experiment 2 are also relevant to the relationship between fluency and attitudes. When transcription accuracy was reduced by noise for the baseline speaker (Figure 3.2), speaker evaluations were not negatively affected, though intelligibility ratings were (Figures 3.2 and 3.3). For the target speaker, noise was associated with more negative evaluations of warmth, competence, and closeness for the non-native (test) guise, but not the native (control) guise (Figure 3.8). These results are consistent with a naïve theory of speech processing fluency discussed in Chapter 1, by which listeners attribute disfluency to the speaker (as opposed to another source of disfluency, such as noise) more for non-native than native accents.

The absence of an effect of objective fluency on both subjective fluency and speaker evaluations indicates a lack of evidence for a causal relationship between fluency and attitudes, inconsistent with the model hypothesized in Chapter 1 (Figure 1.1). The processing fluency research consulted in developing that model (see section 1.2.1) indicates that fluency can influence attitudes both by the fundamental positive valence of cognitive ease and the theory-driven interpretation of that ease (see Forster et al., 2016; Winkielman et al., 2003). The limited research that has examined objective fluency, subjective fluency, and evaluation together (Forster et al., 2013; von Helversen et al., 2008) has found evidence that objective and subjective fluency are positively correlated, that subjective ease is reliably correlated with positive evaluations, and that objective ease appears to only sometimes correlate with positive evaluation, depending on the source of ease (i.e., evidence for the correlation has been found only when fluency was manipulated by the presentation duration of stimuli). The lack of evidence in the present research for a direct effect of objective processing fluency on attitudes is not inconsistent with that research, but the evidence of a correlation between objective and subjective fluency is more limited than expected. The tenuous correlation of objective and subjective fluency, the differential effects of noise-caused disfluency on attitudes toward the non-native and native guises, and the correlation of subjective fluency with social closeness suggest that the theory-driven interpretation of ease in speech perception may interact with stereotypes or expectations about non-native speech (which were not measured here) that influence both the subjective perception of fluency and attitudes toward the speaker. This possibility will be further discussed in the next section and is included in the revised model of processing fluency proposed in section 4.3.

4.2.3 Changes in Processing Fluency and Speaker Evaluation Over Time

Notwithstanding the pattern of results implicating listener expectations in subjective fluency and attitude judgments, the cognitive effort caused by non-native speech may still have had an effect on speaker evaluations. However, if participants experienced difficulty comprehending the non-native accent because they expected comprehension to be difficult, the increased ease that comes with experience may not have been sufficient to change listeners' perception of effort and their attitudes toward the speaker. This section reviews the evidence from these two experiments for a change over time in participants' speaker evaluation on average, then turns to the relationship between changes in fluency and attitudes for individual participants.

Neither experiment provided evidence for an improvement in attitude toward the non-native guise over time. This outcome may seem, at first, consistent with the transcription accuracy results, which do not show evidence of listener adaptation to the non-native accent in either experiment. That is, if adaptation did not occur as expected, the hypothesized improvement in attitudes would not be expected either. My assessment of the full pattern of results, however, is that the lack of improvement in attitude is consistent with the interpretation developed in the preceding sections, that attitudes (along with subjective fluency) are not directly related to objective fluency in the present research.

First, pupil dilation and transcription time results from Experiment 1 did show the predicted evidence that listeners adapted to the non-native accent, while transcription accuracy was near ceiling and did not change over the experiment. In Experiment 2, transcription accuracy improved on average for all participants. Attitude ratings, on the other hand, remained unchanged or worsened over time. The dimension (i.e., warmth, competence, or closeness) on which ratings did not change or decreased differed somewhat across experiments and participant groups (non-native or native guise), but in no case did attitudes improve. Thus, the pattern is that attitudes became more negative on one or more dimensions even while accuracy remained unchanged (Experiment 1) or improved over time (Experiment 2). This is consistent with the evidence discussed in section 4.2.2 showing that attitudes toward the speaker were not related to objective measures of fluency but were related to subjective fluency. It is not clear why attitudes toward the speaker may have worsened over time. As with subjective fluency, the change may have been related to negative affect induced by fatigue or boredom from the experimental task or due to changes in social judgment as participants gained information about the speaker.

For participants individually, change in speaker evaluation showed no relationship with change in objective fluency but was positively related to change in subjective fluency. Evidence for this relationship was most clear for closeness ratings in both experiments (Figures 2.9, 3.10, and 3.11). In Experiment 2, there was some evidence for an effect of intelligibility change on warmth and competence ratings (Figures 3.10 and 3.11), which was not observed in Experiment 1 perhaps due to the absence of controls for individual participants' interpretation of the warmth and competence adjectives. Overall, the effects of intelligibility change on attitude change were weak or moderate, with most effect estimates lower than 0.4, and it was common for some participants in all conditions to show a decrease in intelligibility rating and an improvement in attitude or an increase in intelligibility rating and a decline in attitude. It is not clear, from the data collected, what contributed to individual variation in changes in speaker evaluation in the domains of intelligibility, warmth and competence attributes, and social closeness. To explore the possibility that substantial exposure to more than one language might explain some of this variation, language history questionnaire responses were examined to identify participants with multilingual home or family environments. Six of 22 test participants in Experiment 1 and 18 of 38 test participants in Experiment 2 were identified as belonging to this category. Comparisons of changes in Intelligibility, Warmth, Competence, and Closeness ratings between these participants and those who reported monolingual English home environments revealed no significant differences. These results should be interpreted with abundant caution, as neither experiment was designed to test this difference. Future research examining objective fluency, subjective fluency, adaptation, and attitudes may benefit from measuring cognitive, personality, or social identity variation among participants.

4.3 A Revised Model of Processing Fluency and Language Attitudes

The present findings indicate that the model of processing fluency and language attitudes proposed in Chapter 1 (Figure 1.1) is incomplete. A revised model, based on the results of Experiments 1 and 2, is shown in Figure 4.2. In this model, the effect of objective processing effort on subjective processing fluency and attitudes is mediated by naïve theories (a), which are likely influenced by listeners' stereotypes and language ideologies (b). Listeners' beliefs and the theory-driven interpretation of fluency both influence judgments based on the inferred social category of the speaker (c), which the present results indicate affect both the subjective perception of fluency and attitudes toward the speaker (d). While adaptation to an accent with experience leads to objectively less effortful processing (e), the direct effect of objective fluency on subjective fluency is only tentatively supported by the results of Experiment 2, and is indicated by a dashed arrow (f). Similarly, the potential for change in subjective processing fluency and attitudes is not clearly supported, though some participants did show improvement for both over time. (Subjective fluency and attitudes may also deteriorate over time, as was the case for control participants listening to a native accent.) This potential is speculatively related to the flexibility of listeners' social judgments, as indicated by dashed lines (g). While this model is consistent with the results of Experiments 1 and 2, additional research is needed to clarify how processing fluency affects attitudes in the domain of language. The next section presents some of the outstanding questions and possible avenues for future research.

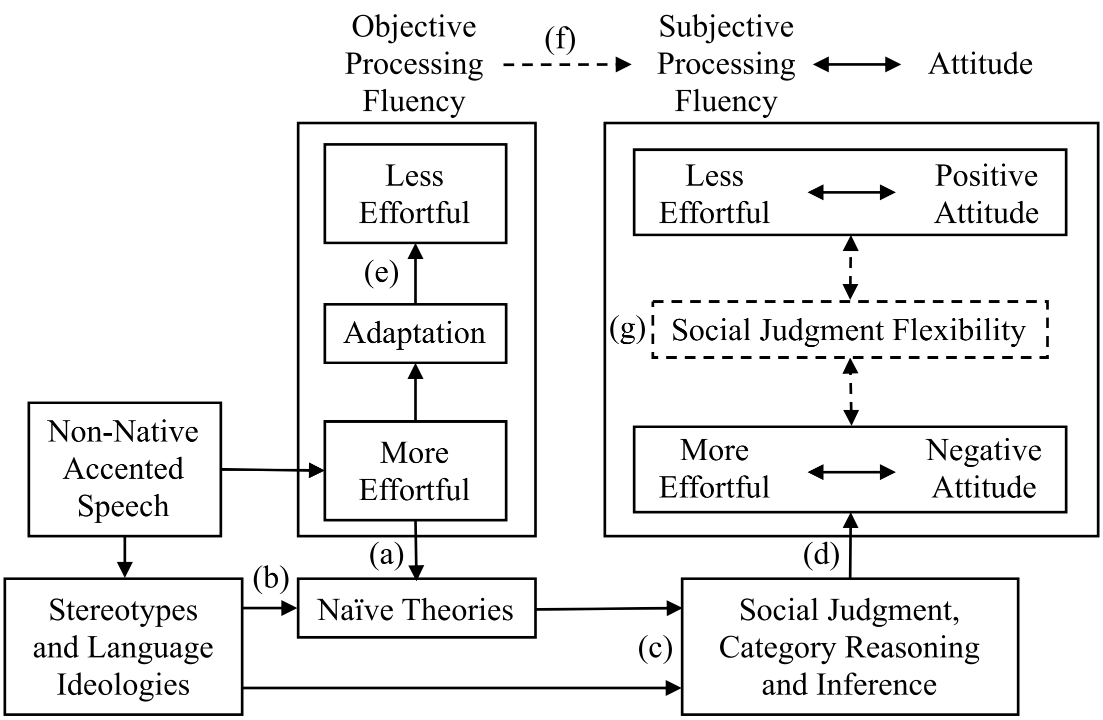

Figure 4.2 A revised model of the relationship between objective processing fluency, subjective processing fluency, and attitude, based on the results of Experiments 1 and 2. (a)-(g) refer to discussion points in the text.

4.4 Directions for Future Research

The primary question left largely unanswered by this dissertation is how naïve theories and expectations about difficult comprehension and its attribution to the speaker or context interact with objective fluency (as measured by, e.g., physiological response, accuracy, or response time) to influence listeners' subjective perception of fluency and attitudes toward the speaker ([a]-[d] in Figure 4.2). The perception of fluency is likely strongly influenced by naïve theories specific to the linguistic domain about what types of speech and speakers are easy or difficult to understand. Previous research has focused on the links between subjective fluency and speaker evaluation without measuring objective fluency (Dragojevic et al., 2017; Dragojevic & Giles, 2016; Lev-Ari & Keysar, 2010). As a result it may be that the results of this research do not indicate the role disfluency plays in negative language attitudes, but rather the strong relationship between the perception of processing ease and attitudes toward the speaker that follow from listener's language ideologies, naïve theories, or expectations about non-native speakers. Future research is necessary to identify the contexts in which objective fluency is reflected in participants' conscious experience and what factors may influence the relationship between objective and subjective fluency ([f] in Figure 4.2).

The effect of expectations on the perception of fluency has important implications for improving attitudes toward non-native speakers. To the extent that disfluency has a negative effect on attitudes, increasing processing fluency holds promise for countering those negative evaluations. However, a fluency-driven model of language attitudes only predicts change if listeners are aware of the increase in fluency. The effects of learned expectations or language ideologies may be strong and require specific interventions to make listeners aware of the reality of their metacognitive experience of processing ease. Expectations based on other social information, such the speaker's appearance, may also affect processing fluency. McGowan (2015) showed a photograph identified as the speaker to participants and asked them to transcribe sentences spoken in Chinese-accented English. Participants who had seen a photograph of an Asian face were more accurate in their transcriptions than participants who had seen a Caucasian face. This suggests that social information like stereotype-based expectations may be involved in linguistic processing, facilitating processing when expectations are upheld, and may influence processing fluency. A mismatch between stereotype-driven expectations and experience may also prompt conceptual disfluency as speakers reevaluate their categorization of the speaker. The potentially interactive nature of stereotypes and processing fluency suggests that challenging language-related stereotypes may also improve perceived fluency in comprehending an unfamiliar accent.

Some of these outstanding questions might be addressed in a more naturalistic setting. For example, universities in the United States frequently employ non-native speakers of English as undergraduate instructors, and attitudes toward these instructors have been a common focus of language attitudes research (e.g., Kang et al., 2014; Rubin, 1992). Research in a classroom setting would provide listeners with repeated interaction with a non-native speaker, which may affect how they integrate social information and processing fluency in forming an attitude toward the speaker. For example, after experience over time, some student-listeners may perceive the instructor more as an individual and rely less on social category inferences. If the subjective perception of fluency is affected by stereotypes about non-native speakers as a group, such individuation may speculatively increase the correlation between subjective fluency and objective measures of fluency. A classroom setting would also facilitate investigating the effects of feedback, i.e., information provided to participants about the success or failure of communication. Implicit feedback, perhaps in the form of misunderstanding or the need for repetition, occurs in everyday communication and might be measured or manipulated in the classroom to understand how such feedback is integrated into listeners' perception of ease or effort. Explicit feedback, like assignment feedback and grades, may also inform perceptions of fluency in such a setting (though understanding the potentially separate effects of subject-matter comprehension and speech comprehension may present methodological challenges). Investigating these and other features of everyday communication may clarify the relationship between ease of comprehension and listeners' perception of ease in ways that the present results do not.

Research on the relationship between objective fluency and listeners' experience may be complemented by research on language ideologies related to non-native accents. Language attitudes research has identified ideologies by which listeners ignore or deny their own role in successful communication, and these ideologies play a part in discrimination against speakers with unfamiliar or difficult accents (Lippi-Green, 2012, pp. 93–96, 149). Building on this research, the present results suggest that ideologies about communication experience may also play a role in attitudes toward non-native speakers. That is, building on what is known about what language users believe about how speakers should speak or how listeners should adapt, future research may explore ideologies about when and whether communication should be easy or difficult. In everyday application, rather than countering the assumption that speakers bear responsibility for successful communication by explaining that listeners play a role in communication, it may be worthwhile to challenge the assumption that communication should be easy with the present finding that difficult communication becomes less effortful with experience.

Future research may also benefit from different methodologies than the ones used here. This dissertation measured attitude change for participants in real time, using adjectives chosen to target particular dimensions of speaker evaluation and presenting different adjectives at the beginning, middle, and end of listening and transcription task to measure changes brought about by perceptual adaptation to the speaker's accent. While this approach has the benefit of reflecting how attitude change due to perceptual adaptation may appear in everyday experience with unfamiliar accents, it is difficult to measure attitudes in such a way that participants are not aware of it. Much of the research on attitude change (see Bodenhausen & Gawronski, 2013 for a review) instead uses an apparent-time approach, in which one group of participants receives the attitude-change "treatment" and the control group does not. Future research investigating how processing fluency and attitudes change depending on factors like adaptation, expectations, and context may benefit from using such a design.

4.5 Conclusion

This dissertation showed evidence that non-native accents can make processing more effortful for native listeners and that cognitive effort decreases over time as listeners adapt, but it showed only tentative evidence that objective measures of fluency were reflected in listeners' self-reported ratings of ease or effort in comprehending the speaker. The subjective perception of processing ease was associated with more positive evaluations of the speaker on the dimensions of warmth, competence, and closeness, but none of these judgments, on average, improved over time. These results indicate that caution may be warranted in attributing negative attitudes toward non-native speakers to disfluency in speech processing. While disfluency likely plays a role in the perception of non-native speech and speakers, it appears to interact with expectations about unfamiliar speech, which may be learned through experience or culturally-salient stereotypes.

For non-native speakers and their interlocutors who experience difficulty in communication, the present results are not without hope. Comprehension of an unfamiliar accent is likely to get easier in a short period of time as a listener's cognitive systems adapt to the new speech patterns. The more a listener recognizes this increased ease, the more that ease may lead to improved attitudes toward the speaker.

APPENDIX A

Transcription Task Sentence Stimuli

Predictable Context	**Unpredictable Context**
A bicycle has two wheels.	She read about his lounge.
A book tells a story.	This is his favorite level.
A chair has four legs.	She looked at his test.
A pigeon is a kind of bird.	The groups have talked about my pool.
A rose is a type of flower.	The class will look at a fellow.
A wristwatch is used to tell the time.	The old men won't talk about the year.
After my bath I dried off with a towel.	The old woman hopes they asked about a buggy.
An orange is a type of fruit.	The children want to ask about sound.
At breakfast he drank some orange juice.	The person hasn't called about our pitch.
Birds build their nests in trees.	Dad would point at our snail.
Bob wore a watch on his wrist.	They are glad she discussed the bowls.
Cut the meat into small pieces.	Mom thinks about the ticket.
Elephants are big animals.	He speaks about alcohol.
February has twenty eight days.	The group might read about love.
Football is a dangerous sport.	This is her favorite plant.
For dessert he had apple pie.	The child had considered their cow.
For your birthday I baked a cake.	The children shouldn't know about our bill.
I wear my hat on my head.	We don't want to consider their town.
I wrote my name on a piece of paper.	The groups would have a problem with order.
In spring the plants are full of green leaves.	The boy was glad I pointed at a drawer.
In the morning it gets light and in the evening it gets dark.	The short man doesn't want the girl to think that it is pink.
Last night they had beef for dinner.	Dad hears they spoke about my answer.
Many people like to start the day with a cup of coffee.	The whole group was glad mom had a problem with her issues.
Monday is the first day of the week.	The men hope I heard about his work.

Predictable Context	Unpredictable Context
My clock was wrong so I got to school late.	The classes want us to know that mom was sick.
My parents sister and I are a family.	The children want to hear about a policy.
People wear gloves on their hands.	I can call about its parts.
People wear shoes on their feet.	She thinks that it is mild.
Rain falls from clouds in the sky.	Everybody shouldn't have pointed at the can.
She cut the cake with a knife.	I wouldn't have thought about the quilt.
She laid the meal on the table.	We hoped he was interested in her party.
She looked at herself in her mirror.	I shouldn't have called about their ladder.
She made the bed with clean sheets.	The group can't know about the fridge.
Sugar tastes very sweet.	It can't be brown.
The boy laughed because the joke was very funny.	The child didn't have a problem with our style.
The bread was made from whole wheat.	The groups could ask about his lease.
The child dropped the dish and it broke.	The child hoped he knew about their room.
The color of a lemon is yellow.	The girls didn't think it was narrow.
The heavy rains caused a flood.	Everyone won't have considered the swing.
The lady wears earrings in her ears.	The boy might talk about her duck.
The meat from a pig is called pork.	The class won't have asked about the barn.
The pan that was just in the oven is very hot.	The old women don't want him to hear it is red.
The picture is hung high on the bedroom wall.	The boys wouldn't have been interested in a game.
The sick woman went to see a doctor.	They can't have had a problem with my letter.
The sport shirt has short sleeves.	We had heard about her prayers.
The stars come out at night.	The class knew about my house.
The team was trained by their coach.	The man shouldn't read about our guide.
The war plane dropped a bomb.	They are glad she discussed the bowls.
To cool her drink she added a few cubes of ice.	The tall woman is glad she was interested in their aid.
We heard the ticking of the clock.	The girls couldn't speak about our claim.

Table A.1 Transcription task sentence stimuli

APPENDIX B

Language History and Debriefing Questionnaire

1. Age:
2. Gender:
3. Country of birth:
4. What is your native language? (If you consider yourself a native speaker of more than one language, please specify.)
5. List all languages other than English that you know. Please indicate your age when you began learning each language, and rate your overall ability in each language now.
6. What languages were spoken in the home, including with friends and relatives, while you were growing up?
7. If you have lived or travelled in countries where English is not the predominant language, please indicate the name(s) of the country or countries, your length of stay, and the language(s) you learned or tried to learn.
8. If there is anything else that you feel is interesting or important about your language background or language use, please comment below.
9. Thinking back to the listening task, were the questions about the speaker reasonable, or did any descriptions stand out to you?
10. Were you trying to figure out anything about the speaker? (For example, where he was from.)
11. Did you wonder about the specific research question this study was intended to answer? What do you think the question was?

APPENDIX C

Transcribed Non-Words and Assigned Pronunciations

Transcription	Pronunciation	Transcription	Pronunciation
acholol	[ækəlɑl]	hig	[hɪg]
apeasian	[əpiʒən]	hos	[hɑs]
apress	[əprɛs]	leds	[lɛdz]
brung	[bɹəŋ]	lounde	[laʊnd]
can't've	[kæntəv]	peeshian	[piʃən]
childed	[tʃaɪldəd]	punted	[pʌntəd]
couldn	[kʊdən]	shild	[ʃɪld]
cril	[kɹɪl]	shoudn	[ʃʊdən]
dears	[dɪɹz]	shouldn't've	[ʃʊdəntəv]
ded	[dɛd]	spects	[spɛkts]
drawar	[dɹɔɹ]	tickin	[tɪkɪn]
drower	[dɹɔɹ]	traineded	[tɹeɪndəd]
epision	[əpɪʒən]	tranded	[tɹændəd]
glatsche	[glɑtʃə]	trialed	[tɹaɪld]
gril	[gɹɪl]	untipped	[əntɪpt]
grot	[gɹɑt]	waman	[wɑmən]
hasn	[hæzən]	wearin'	[wɛɹɪn]
hdac	[hədæk]	wonem	[wɪnəm]

Table C.1 Transcribed non-words and assigned pronunciations.

BIBLIOGRAPHY

Adank, P., & Janse, E. (2010). Comprehension of a novel accent by young and older listeners. *Psychology and Aging*, *25*(3), 736–740. https://doi.org/10.1037/a0020054

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the Tribes of Fluency to Form a Metacognitive Nation. *Personality and Social Psychology Review*, *13*(3), 219–235. https://doi.org/10.1177/1088868309341564

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America*, *133*(3), EL174–EL180. https://doi.org/10.1121/1.4789864

Banks, B., Gowen, E., Munro, K. J., & Adank, P. (2015). Cognitive predictors of perceptual adaptation to accented speech. *The Journal of the Acoustical Society of America*, *137*(4), 2015–2024. https://doi.org/10.1121/1.4916265

Bastian, B., Lusher, D., & Ata, A. (2012). Contact, evaluation and social distance: Differentiating majority and minority effects. *International Journal of Intercultural Relations*, *36*(1), 100–107. https://doi.org/10.1016/j.ijintrel.2011.02.005

Bergsieker, H. B., Leslie, L. M., Constantine, V. S., & Fiske, S. T. (2012). Stereotyping by omission: Eliminate the negative, accentuate the positive. *Journal of Personality and Social Psychology*, *102*(6), 1214–1238. https://doi.org/10.1037/a0027717

BNC Consortium. (2007). *British National Corpus* (3 (BNC XML Edition)) [Computer software]. http://www.natcorp.ox.ac.uk/

Bodenhausen, G. V., & Gawronski, B. (2013). Attitude Change. In D. Reisberg (Ed.), *The Oxford Handbook of Cognitive Psychology* (pp. 957–969). Oxford University Press.

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America*, *121*(4), 2339–2349. https://doi.org/10.1121/1.2642103

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729. https://doi.org/10.1016/j.cognition.2007.04.005

Briñol, P., Petty, R. E., & Tormala, Z. L. (2006). The Malleable Meaning of Subjective Ease. *Psychological Science*, *17*(3), 200–206. https://doi.org/10.1111/j.1467-9280.2006.01686.x

Carnegie Mellon University. (2014). *Carnegie Mellon Pronouncing Dictionary* (0.7b) [Computer software]. http://svn.code.sf.net/p/cmusphinx/code/trunk/cmudict/

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, *116*(6), 3647–3658. https://doi.org/10.1121/1.1815131

Derwing, T. M. (2003). What Do ESL Students Say About Their Accents? *Canadian Modern Language Review*, *59*(4), 547–566. https://doi.org/10.3138/cmlr.59.4.547

Derwing, T. M., Rossiter, M. J., & Munro, M. J. (2002). Teaching Native Speakers to Listen to Foreign-accented Speech. *Journal of Multilingual and Multicultural Development*, *23*(4), 245–259. https://doi.org/10.1080/01434630208666468

Dragojevic, M., & Giles, H. (2016). I Don't Like You Because You're Hard to Understand: The Role of Processing Fluency in the Language Attitudes Process. *Human Communication Research*, *42*(3), 396–420. https://doi.org/10.1111/hcre.12079

Dragojevic, M., Giles, H., Beck, A.-C., & Tatum, N. T. (2017). The fluency principle: Why foreign accent strength negatively biases language attitudes. *Communication Monographs*, *84*(3), 385–405. https://doi.org/10.1080/03637751.2017.1322213

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Forster, M., Leder, H., & Ansorge, U. (2013). It felt fluent, and I liked it: Subjective feeling of fluency rather than objective fluency determines liking. *Emotion*, *13*(2), 280–289. https://doi.org/10.1037/a0030115

Forster, M., Leder, H., & Ansorge, U. (2016). Exploring the subjective feeling of fluency. *Experimental Psychology*, *63*(1), 45–58. https://doi.org/10.1027/1618-3169/a000311

Gluszek, A., & Dovidio, J. F. (2010a). The Way They Speak: A Social Psychological Perspective on the Stigma of Nonnative Accents in Communication. *Personality and Social Psychology Review*, *14*(2), 214–237. https://doi.org/10.1177/1088868309359288

Gluszek, A., & Dovidio, J. F. (2010b). Speaking With a Nonnative Accent: Perceptions of Bias, Communication Difficulties, and Belonging in the United States. *Journal of Language and Social Psychology*, *29*(2), 224–234. https://doi.org/10.1177/0261927X09359590

Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, *33*(4), 457–461. https://doi.org/10.1111/j.1469-8986.1996.tb01071.x

Heaton, H., & Nygaard, L. C. (2011). Charm or Harm: Effect of Passage Content on Listener Attitudes Toward American English Accents. *Journal of Language and Social Psychology*, *30*(2), 202–211. https://doi.org/10.1177/0261927X10397288

Huyck, J. J., & Johnsrude, I. S. (2012). Rapid perceptual learning of noise-vocoded speech requires attention. *The Journal of the Acoustical Society of America*, *131*(3), EL236–EL242. https://doi.org/10.1121/1.3685511

Kang, O., & Rubin, D. L. (2009). Reverse Linguistic Stereotyping: Measuring the Effect of Listener Expectations on Speech Evaluation. *Journal of Language and Social Psychology*, *28*(4), 441–456. https://doi.org/10.1177/0261927X09341950

Kang, O., Rubin, D., & Lindemann, S. (2014). Mitigating U.S. Undergraduates' Attitudes Toward International Teaching Assistants. *TESOL Quarterly*, *49*(4), 681–706. https://doi.org/10.1002/tesq.192

Ladefoged, P., & Broadbent, D. E. (1957). Information Conveyed by Vowels. *The Journal of the Acoustical Society of America*, *29*(1), 98–104. https://doi.org/10.1121/1.1908694

Lev-Ari, S. (2015). Comprehending non-native speakers: Theory and evidence for adjustment in manner of processing. *Frontiers in Psychology*, *5*, 1546. https://doi.org/10.3389/fpsyg.2014.01546

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*(6), 1093–1096. https://doi.org/10.1016/j.jesp.2010.05.025

Leyens, J.-P., & Yzerbyt, V. Y. (1992). The ingroup overexclusion effect: Impact of valence and confirmation on stereotypical information search. *European Journal of Social Psychology*, *22*(6), 549–569. https://doi.org/10.1002/ejsp.2420220604

Lindemann, S. (2005). Who speaks "broken English"? US undergraduates' perceptions of non-native English1. *International Journal of Applied Linguistics*, *15*(2), 187–212. https://doi.org/10.1111/j.1473-4192.2005.00087.x

Lindemann, S., & Subtirelu, N. (2013). Reliably Biased: The Role of Listener Expectation in the Perception of Second Language Speech. *Language Learning*, *63*(3), 567–594. https://doi.org/10.1111/lang.12014

Lippi-Green, R. (2012). *English with an accent: Language, ideology and discrimination in the United States* (2nd ed). Routledge.

Mandler, G., Nakamura, Y., & Van Zandt, B. J. (1987). Nonspecific effects of exposure on stimuli that cannot be recognized. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(4), 646–648. https://doi.org/10.1037/0278-7393.13.4.646

McGowan, K. B. (2015). Social Expectation Improves Speech Perception in Noise. *Language and Speech*, *58*(4), 502–521. https://doi.org/10.1177/0023830914565191

Munro, M. J., & Derwing, T. M. (1995a). Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. *Language Learning*, *45*(1), 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Munro, M. J., & Derwing, T. M. (1995b). Processing Time, Accent, and Comprehensibility in the Perception of Native and Foreign-Accented Speech. *Language and Speech*, *38*(3), 289–306. https://doi.org/10.1177/002383099503800305

Northwestern University Linguistics Labs. (2015). *LTAS to Sound*. http://groups.linguistics.northwestern.edu/documentation/v2/scripts.html

Open Source Collaboration. (2015). *SoX* (Version 14.4.2) [Computer software]. http://sox.sourceforge.net/Main/HomePage

Oppenheimer, D. M. (2006). Consequences of erudite vernacular utilized irrespective of necessity: Problems with using long words needlessly. *Applied Cognitive Psychology*, *20*(2), 139–156. https://doi.org/10.1002/acp.1178

Pearson, A. R. (2011). *Intergroup Fluency: A Metacognitive Approach to Intergroup Perception* [Ph.D., Yale University]. http://search.proquest.com.proxy.lib.umich.edu/dissertations/docview/922284060/abstract/C113A2835D404635PQ/1

Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and Phonetic Experiments on American English Dialect Identification. *Journal of Language and Social Psychology*, *18*(1), 10–30. https://doi.org/10.1177/0261927X99018001002

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of Perceptual Fluency on Affective Judgments. *Psychological Science*, *9*(1), 45–48. JSTOR.

Reber, R., Wurtz, P., & Zimmermann, T. D. (2004). Exploring "fringe" consciousness: The subjective experience of perceptual fluency and its objective bases. *Consciousness and Cognition*, *13*(1), 47–60. https://doi.org/10.1016/S1053-8100(03)00049-7

Richman, J. E., McAndrew, K. G., Decker, D., & Mullaney, S. C. (2004). An evaluation of pupil size standards used by police officers for detecting drug impairment. *Optometry-Journal of the American Optometric Association*, *75*(3), 175–182.

Rubin, D. L. (1992). Nonlanguage Factors Affecting Undergraduates' Judgments of Nonnative English-Speaking Teaching Assistants. *Research in Higher Education*, *33*(4), 511–531.

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, *71*(6), 1207–1218. https://doi.org/10.3758/APP.71.6.1207

Schwarz, N. (2004). Meta-cognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology*, *14*(4), 332–348.

Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America*, *125*(5), 3306–3316. https://doi.org/10.1121/1.3101452

Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(6), 679–692. https://doi.org/10.1002/wcs.1323

Strand, E. A. (1999). Uncovering the Role of Gender Stereotypes in Speech Perception. *Journal of Language and Social Psychology*, *18*(1), 86–100. https://doi.org/10.1177/0261927X99018001006

Subtirelu, N. C., & Lindemann, S. (2014). Teaching First Language Speakers to Communicate Across Linguistic Difference: Addressing Attitudes, Comprehension, and Strategies. *Applied Linguistics*, 1–20. https://doi.org/10.1093/applin/amu068

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat Corpus of Native-and Foreign-accented English: Communicative Efficiency across Conversational Dyads with Varying Language Alignment Profiles. *Language and Speech*, *53*(4), 510–540. https://doi.org/10.1177/0023830910372495

von Helversen, B., Gendolla, G. H. E., Winkielman, P., & Schmidt, R. E. (2008). Exploring the hardship of ease: Subjective and objective effort in the ease-of-processing paradigm. *Motivation and Emotion*, *32*(1), 1–10. https://doi.org/10.1007/s11031-008-9080-6

Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(6), 1235–1253. https://doi.org/10.1037/0278-7393.19.6.1235

Winkielman, P., & Cacioppo, J. T. (2001). Mind at ease puts a smile on the face: Psychophysiological evidence that processing facilitation elicits positive affect. *Journal of Personality and Social Psychology*, *81*(6), 989–1000. https://doi.org/10.1037//0022-3514.81.6.989

Winkielman, P., Halberstadt, J., Fazendeiro, T., & Catty, S. (2006). Prototypes Are Attractive Because They Are Easy on the Mind. *Psychological Science*, *17*(9), 799–806. https://doi.org/10.1111/j.1467-9280.2006.01785.x

Winkielman, P., Schwarz, N., Fazendeiro, T. A., & Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion.* (2003-04670-007; pp. 189–217). Lawrence Erlbaum Associates Publishers.

Winn, M. B., Edwards, J. R., & Litovsky, R. Y. (2015). The Impact of Auditory Spectral Resolution on Listening Effort Revealed by Pupil Dilation: *Ear and Hearing*, *36*(4), e153–e165. https://doi.org/10.1097/AUD.0000000000000145

Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get Started. *Trends in Hearing*, *22*. https://doi.org/10.1177/2331216518800869

Yzerbyt, V. Y., & Leyens, J.-P. (1991). Requesting information to form an impression: The influence of valence and confirmatory status. *Journal of Experimental Social Psychology*, *27*(4), 337–356. https://doi.org/10.1016/0022-1031(91)90030-A

Zekveld, A. A., Koelewijn, T., & Kramer, S. E. (2018). The Pupil Dilation Response to Auditory Stimuli: Current State of Knowledge. *Trends in Hearing*, *22*. https://doi.org/10.1177/2331216518777174