Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

______________________________ _______________

Eva Michelle Lewandowski March 20, 2018

The automaticity of entrainment:
Perceptual, social, and individual differences in vocal alignment

By

Eva Michelle Lewandowski
Doctor of Philosophy

Psychology

Lynne C. Nygaard, Ph.D.
Advisor

Harold Gouzoules, Ph.D.
Committee Member

Stella F. Lourenco, Ph.D.
Committee Member

Donald Tuten, Ph.D.
Committee Member

Phillip Wolff, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

The automaticity of entrainment:
Perceptual, social, and individual differences in vocal alignment

By

Eva Michelle Lewandowski
M.A., Emory University, 2014

Advisor: Lynne C. Nygaard, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Psychology
2018

Abstract

The automaticity of entrainment:
Perceptual, social, and individual differences in vocal alignment
By Eva M. Lewandowski

Vocal alignment, or entrainment to the speech characteristics of others, has been hypothesized to strengthen social affiliation and enhance efficiency of spoken communication. This dissertation investigates vocal alignment in terms of the automaticity of its component processes. Two experiments test the hypothesis that alignment is driven primarily by automatic, relatively resource-free perceptual-motor processing, which can be modulated by intentional, resource demanding cognitive control mechanisms. Experiment 1 focused on the relative automaticity of perceptual-motor processes and explored the extent to which vocal alignment is “efficient” (e.g., does not require attentional or cognitive resources) by introducing a working memory load. Experiment 2 examined the social and cognitive control processes associated with vocal alignment by pairing targets of alignment (e.g., model talkers) with socially positive or negative information. This was done to determine if trait inferences influenced alignment behavior and whether social evaluation required cognitive resources. Both experiments also explored individual differences in the tendency to entrain, focusing both on differences in dispositional characteristics and productive flexibility. The findings of Experiment 1 suggest that vocal alignment may not be necessarily resource demanding. Experiment 2 demonstrates that using social assessments to guide vocal alignment, however, may require cognitive resources. Certain types of flexibility in speech production (e.g., variation in vowel spectra) were associated with differences in vocal alignment, but individual differences in empathy were not. These results inform the interface between language and other socio-cognitive systems, and provide a basis for further exploration of entrainment in speech behavior.

The automaticity of entrainment:
Perceptual, social, and individual differences in vocal alignment

By

Eva Michelle Lewandowski
M.A., Emory University, 2014

Advisor: Lynne C. Nygaard, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Psychology
2018

Acknowledgements

As my dissertation committee member Phil Wolff once said, graduate school is like running a marathon. There are indeed parallels: in order to make it across the finish line, one must pace oneself; and around mile 18, one's legs start to burn. The marathon comparison is also apt because it takes a team to prepare the runner to cross the finish line, from volunteers at water stations to first responders to the cheering crowd. As a graduate student, I have had the incredible fortune to be part of the lives of many talented people who have helped me approach the finish line.

First, I would like to acknowledge my family for their support and the importance they placed on my education. I would never have made it to Emory if not for my mother making sure I always had the time and resources I needed for my studies, for making my quality of education her highest priority; if not for my father's insistence that my job was to always absorb as much knowledge as I could. I also want to acknowledge my fiancé, Adam Segal, for believing in me enough to follow me to Atlanta six years ago. Whenever the road through graduate school seemed impossible to travel, he was always there to lift my spirits and help me move forward. I would not have made it as far as I have without my family's unwavering support.

I am also exceptionally lucky for the friends and colleagues I have met at Emory. I thank my colleagues in the Speech and Language Perception Lab, especially Kelly McCormick and Christina Tzeng, for being scholarly examples and for their input on my research ideas. I must also thank the research assistants who contributed to my projects: Julia Abovich, Taylor Block, Isabella Detchmendy, Milly Huizinga, Ashwashi Menon, Julia Nadel, Trulacia Powell, Aileen Rivell, Tera Robison, and Gabrielle Stravach. I particularly thank Tera Robison, for her scientific curiosity, exceptional work ethic, and reassuring optimism during the longest mile. Without the help of these talented scholars, I simply could not have completed my dissertation. To my dear friends and (past) fellow graduate students, I likewise owe a debt of gratitude. I wish to thank Samuel Hunley, Joanna Berg, Ashley Prichard, Charlie Ferris, Robert Thorstad, Jay Schwartz, Christina Rogers, and Marianne Florian for their intellectual and social companionship. I also thank Bree Beal for permitting me to use his character narratives in my dissertation. His generosity spared me countless hours. Whether I needed to play board games or test the logic of an argument, to commiserate or celebrate, I always had people I could count on.

Of all the people at Emory who have shaped me into a better scholar, I must extend special gratitude to my mentor, Lynne Nygaard, for her invaluable counsel throughout the years. She has been a conduit of knowledge, encouraging me to pursue meaningful research questions and challenging me to expand my intellectual horizons. By encouraging me to participate in the Center for Mind, Brain, and Culture's certificate program, she has helped me engage with speech perception and speech production in a broader context of society and behavior. Lynne has supported me in all aspects of my research, from wrestling with theoretical questions to developing clear and precise study design to considering broader implications of my work. She has faced down tight deadlines alongside me, advocated for me, and been an inspiration and a role model. Lynne is a paragon of mentorship, and her humane guidance has helped me grow as a scholar and as a person. The lessons I have learned from her in thought, writing, and scholarship will stay with me forever.

Table of Contents

List of Tables

List of Equations

List of Figures

The automaticity of entrainment:
Perceptual, social, and individual differences in vocal alignment

Consider the following linguistic exchange: a person enters a shop and converses with the cashier. The cashier has a strong regional dialect and, by contrast, the customer speaks a standard dialect. As the pair converses, they each subtly adopt characteristics of the other's speaking style. This hypothetical case is an illustration of *vocal alignment*, a process by which language users entrain to the speech characteristics of a heard other. Language users can align to the structure of speech along multiple linguistic levels, including phonetic detail (Abrego-Collier, Grove, Sonderegger, & Yu, 2011), lexical choice (Branigan et al., 2011), and syntactic structure (Branigan, Pickering, & Cleland, 2000). Both children and adults engage in linguistic entrainment (Goldinger, 1998; Nielson, 2014; Subiaul, Winters, Krumpak, & Core, 2016), and similar propensities have been documented in social non-human animals (Balsby & Scarl, 2008; Sugiura, 1998; Zürcher & Burkart, 2017). Research suggests that the tendency to adopt the speaking style of another individual may serve to strengthen social bonds and may result in shared speech and language representations (Dragojevic & Giles, 2016; Giles, Coupland, & Coupland, 1991; Gregory, Dagan, & Webster, 1997; Manson, Bryant, Geravis, & Kline, 2013; Pickering & Garrod, 2004).

The mechanisms responsible for vocal alignment have traditionally been understood as stemming either from social and motivational causes (Giles et al., 1991, Cutler, 2010) or from perception-production links intrinsic to the hearer (Honorof, Weihing, & Fowler, 2011). In this dissertation, I begin by reviewing vocal alignment and theoretical accounts for this behavior. Next, I propose a reframing of this evidence in terms of controlled and automatic behaviors (e.g., Bargh, 1994; Schneider & Schiffrin, 1977) in order to reconcile diverse findings on vocal alignment. I then report two experiments that examine the efficiency, intentionality, and control of vocal alignment behavior. Experiment 1 directly tests the extent to which vocal alignment requires cognitive resources. Experiment 2 explores whether social evaluation modulates alignment behavior and if so, whether evaluation acts in an automatic or resource-demanding way. The data from both experiments are pooled for an analysis of the influence of individual differences in empathy and vocal flexibility on vocal alignment behavior. Finally, the implications of the findings are discussed in the broader context of vocal and behavioral alignment and in terms of the interplay between perceptual-motor and cognitive control factors in language behavior more generally.

Alignment for Social Fitness

To take a broad view of alignment behavior across individuals and species, the success of animals living in social groups depends not only on their physical or mental fitness, but also on their social fitness. That is, the social bonds, alliances, and status an individual achieves have direct impacts for reproductive success (for a recent review, see Silk, 2014). Situated in this broader context, comparative evidence suggests that behavioral alignment may contribute to the social survival of group-living individuals (Balsby & Scarl, 2008; Chartrand & Bargh, 1999; Chartrand & Lakin, 2013; Rukstalis, Fite, & French, 2003; Smolker & Pepper, 1999; Snowdon & Elowson, 1999). For example, orange-fronted conures, dolphins, and spider monkeys live in fission-fusion societies, where group composition changes over time. Contact and alignment to vocal call structure occurs rapidly and may be important for group cohesion (Balsby & Adams, 2011; Bradbury, Cortopassi, & Clemmons, 2001; Ramos-Fernandez, 2005; Janik, 2000). This may be the case for both human and non-human animals.

Shared vocal signals are one way to demarcate social groups and to achieve group cohesion. In humans, shared linguistic dialect is an important group marker. As one example, Heblich, Lameli, and Riener (2015) found that hearing the dialect of an out-group member made participants more likely to choose a competitive over a cooperative strategy in economic games. The importance of dialect is also observed in young children; they prefer to befriend someone of their own dialect group over someone of their own race (Kinzler, Shutts, DeJesus, & Spelke, 2009). Vocal alignment at the level of individuals may be one mechanism by which groups maintain their distinct vocal characteristics, and linguistic changes that accumulate over time may contribute to larger scale language change (Trudgill, 2008; Tuten, 2008).

In non-human animals, vocal signaling may also serve as a group marker. For example, there are suggestions that chimpanzees (Mitani, Hunley, & Murdoch, 1999) and marmosets (de la Torre & Snowdon, 2009; Zürcher & Burkart, 2017) have group-specific call structures (i.e., "dialects"). In these species, it has been shown that genetic relatedness does not predict call similarity among individuals whereas group membership does (Crockford, Herbinger, Vigilant, & Boesch, 2004; Lemasson & Hausberger, 2004; Marshall, Wrangham, & Arcadi, 1999). Entrainment to the characteristics of group members' vocal productions may be a proximate cause of such similarities. In line with this possibility, chimpanzees that were relocated from a Dutch zoo to a Scottish one altered the characteristics of their food grunts (Watson et al., 2015a). Importantly, alignment was not bi-directional; only the Dutch chimps showed alignment to the Scottish ones and only after they had been integrated into the group. While the role of convergent affective states cannot be ruled as a potential cause for grunt matching (Watson et al, 2015b, Fischer et al., 2015), this evidence suggests that chimpanzees, our nearest extant cousin, may use alignment to maintain social groups.

In addition to bringing groups together, vocal alignment can increase cohesion between individual group members. Empirical evidence of this can be found across the animal kingdom. In humans, alignment on speaking rate predicts increased liking (Street, 1984) and increases cooperative behavior in a one-shot prisoner's dilemma, an economic game in which defection is the most monetarily advantageous choice and cooperation confers no long-term benefit (Manson et al., 2010). Social non-human animals show similar vocal matching behavior to social partners. For example, mated pygmy marmosets converge on vocal call structures after becoming a pair (Snowdown & Elowson, 1999). Observations of wild dolphins suggest that allied bottlenose dolphins converge on common whistle types (Smolker & Pepper, 1999). Japanese macaques align to the vocal frequency of playback stimuli, and furthermore, show greater convergence toward playbacks of higher ranked individuals (Lemasson, Jubin, Masataka, & Arlet, 2016). The presence of similar behavior patterns in human and non-human animals points to a deep functional significance of sounding like others. What mechanisms achieve the individual and group-level consequences of alignment in vocal behavior? The current dissertation considers this question for the human animal.

Effects of Social and Situational Factors

Extensive research has demonstrated that social information, such as status and attitudes, and situational factors, such as task demands, influence vocal alignment behavior (Bilous & Kraus, 1988; Pardo, 2006; Pardo, Jay, & Krauss, 2010). For example, Gregory and Webster (1996) analyzed conversations from the television show *Larry King Live* and found that show host Larry King aligned more to the speech of his higher status guests than to his lower status guests. Similarly, British English speakers in an interview setting were more likely to align to a speaker of Received Pronunciation (RP), a high-status dialect of British English, than to a speaker of Bristol English, a lower status dialect of British English (Giles, 1973). These types of findings suggest a sensitivity to the status and speech of others and that this sensitivity appears to guide the extent to which individuals vocally align to a conversational partner.

Gender differences in vocal alignment have also been considered at length, as males and females may use different linguistic strategies to accomplish discursive goals (Gravano et al., 2011; Haas, 1979; Newman, Groom, Handelman, & Pennebaker, 2008). Studies vary, however, in their estimates of whether males or females exhibit more vocal alignment (Babel, 2012; Babel, McGuire, Walters & Nicholls, 2014; Bilous and Krauss, 1988; Namy, Nygaard & Sauertieg, 2002; Pardo, Jordan, Mallari, Scanlon, & Lewandowski, 2013b; Pardo et al., 2017), indicating that the effects of gender may be heavily influenced by other factors.

Indeed, the effect of macro-level social features may be modulated by short-term situational factors or goals. In an observational study, Coupland (1984) reported a Welsh-accented travel agent’s use of nonstandard dialect features (e.g., h-dropping, dropping "g" off of "ing" clusters, etc.) as a function of her client's occupational class. The agent shifted her use of non-standard phonological features toward clients regardless of their occupational class, suggesting that short-term communicative goals can also moderate vocal alignment. In a map task (Anderson et al., 1991), in which instruction givers describe a specific path that instruction receivers draw on a blank version of a map, the instruction giver arguably has the 'higher status' role, as they are the arbiter of the map's correctness. When participants change task roles, thus exchanging social roles, vocal alignment followed initial role assignment (Pardo et al., 2013a).

In general, there is ample evidence supporting a role for social factors and particular situational and communicative contexts in vocal alignment. Theoretical approaches such as Communication Accommodation Theory (CAT) propose that the goals of alignment are to manage social distance by increasing, decreasing, or maintaining linguistic distance and to improve interactional efficiency (Dragojevic & Giles, 2016; Giles & Ogay, 2007; Giles, Willemyns, Gallois, & Anderson, 2007; Giles et al., 1991). However, CAT does not specify the details of how these goals are accomplished, whether these goal states are within or outside conscious awareness, or the cognitive and neural mechanisms that would achieve these interactional goals. The proposed functionality of and impact of social factors on vocal alignment do not necessarily entail that the proximate causes must be social. Without an understanding of the perceptual and cognitive mechanisms that contribute to vocal alignment, our ability to explain and clearly predict when and under what circumstances vocal alignment behavior will occur remains incomplete. Specifying the mechanism may also contribute to our understanding of the functional significance of alignment behavior.

Cognitive and Perceptual Mechanisms Supporting Vocal Alignment

Social information does affect the degree and patterning of vocal alignment, yet this behavior has been demonstrated in experimental paradigms that minimize explicit social and communicative goals (Goldinger, 1998). Participants in *naming* or *shadowing* paradigms, which involve repeating (i.e., naming) pre-recorded speech over headphones, demonstrate alignment to many fine-grained characteristics of speech (Brouwer, Mitterer, & Huettig, 2010; Mitterer & Ernestus, 2008; Shockley, Sabadini, & Fowler, 2004; Walker & Campbell-Kibler, 2015). Convergence on acoustic speech characteristics can be observed even when shadowers passively listen to the model talker's speech and read word lists after the model talker's speech is no longer present (Goldinger & Azuma, 2004; Kim, 2011; Kim, 2012). Although even experimental tasks like the naming paradigm are not devoid of socio-cultural information, they minimize the role of explicit social goals and constraints. Because vocal alignment occurs even in the absence of a clear social motivation to align, psycholinguistic accounts of vocal alignment have emphasized the properties of cognitive, perceptual, and motor architecture that drive the behavior (e.g. Pickering & Garrod 2013; Fowler & Galantucci, 2005).

For example, Fowler and Galantucci (2005) posit that vocal alignment may be due to shared underlying representations that are used for both speech perception and speech production (e.g., Liberman & Mattingly, 1985; Prinz, 1990; Skipper, Devlin, & Lametti, 2017). Due to the fundamental coupling between perception and action, input to the speech perception system may necessarily result in altered output. Therefore, hearing distinctive or atypical speech should result in greater or more salient vocal alignment.

Consistent with this proposal, a number of researchers have found that individuals show greater alignment to speaking varieties that are dissimilar from their own (Babel, McGuire, Walters, & Nicholls, 2014; Lewandowski & Nygaard, under review; Walker & Campbell-Kibler, 2015; but see Kim, Horton, & Bradlow, 2011). Further, Adank, Hagoort, and Bekkering (2010) showed that imitating a novel Dutch accent led to significant improvement in participants' ability to comprehend words spoken in the novel accent. The improvement to intelligibility from explicit imitation was greater than the gain from other types of training, suggesting that alignment is useful for understanding non-standard speech. Speech that is more difficult to understand, such as non-words and words that occur infrequently in the language, also tend to receive greater alignment from both children and adults than high-frequency, familiar words (Goldinger, 1998; Kappes, Baumgaertner, Pesche, & Ziegler, 2009; Subiaul et al., 2016). These types of findings highlight the importance of basic perceptual-motor processing in determining the outcomes of alignment behavior and suggest a consequence of that coupling in terms of intelligibility. The cognitive architecture relating perception and action may support vocal alignment for this purpose.

Another alternative to social-motivational accounts was proposed by Pickering and Garrod (2004; 2006), who advance that priming across multiple levels of linguistic representation results in vocal alignment. The Interactive Alignment Model (IAM) approach is distinct from the perceptual-motor accounts proposed by Fowler and others because they assert alignment is based on general priming of linguistic representations across dyads, rather than exclusively on a fundamental perception-production link. Further, aligning to linguistic construction at one level (e.g., syntax, lexicon) is hypothesized to necessarily lead to synchrony at multiple levels. Primed representations across dyads ensure that speakers have an aligned or synchronized understanding of who, what, when, where, why, and how (e.g., situation representations) of a topic under discussion, which in turn are hypothesized to lead to improved comprehension (Pickering & Garrod, 2006; 2013). In support of the communicative benefits of aligning a situation model, partners in a dyadic task (i.e., find four differences in two highly similar pictures) who showed greatest alignment to one another also tended to show greatest task efficiency (Kim et al., 2011; van Engen et al., 2010). Priming has also been demonstrated across multiple levels of linguistic representation. The well-studied syntactic priming effect, which shows that individuals tend to produce the grammatical constructions to which they have been recently exposed, can be interpreted as another example of priming-based alignment (Bock, 1986; Pickering & Branigan, 1998). Branigan et al. (2000) reported that participants in a picture-matching task were significantly more likely to use a disfavored object label (i.e., "coach" as a disfavored label for "bus") after their task partner had used that label. Extending this work, Postma-Nilsenová, Mol, and Kamoen (2013) found that alignment on syntax increased the likelihood of alignment to acoustic-phonetic properties of speech. These data support the role for general priming in conversational alignment and suggest that it spreads to other levels of linguistic representation, facilitating overall communicative efficiency.

Cognitive and perceptual approaches to speech entrainment can account for alignment behavior but face difficulty in explaining vocal divergence (becoming less similar to an interlocutor). If vocal alignment is driven by priming or perceptual-motor mechanisms that facilitate comprehension across levels of language, then interlocutors should always engage in vocal alignment. Although there are many more examples of alignment in the literature than of divergence (even to targets associated with negative attitudes; Kim, 2012; Lewandowski & Nygaard, under review), it is demonstrably not the case that language users *only* engage in alignment behavior (e.g., Babel, 2010; Babel, 2012; Bourhis & Giles, 1977). Observations of divergence suggest that social and motivational factors must interact with lower level perceptual or cognitive mechanisms in complex ways.

A Unifying Framework

The influence of higher-level social and situational factors on speech behavior does not exclude the possibility that lower-level perceptual or priming processes play causal roles in alignment behavior. In some ways, the separate lines of research in the domain of vocal alignment could reflect the contributions of fundamentally different types of mechanisms to the same phenomenon (Ruch, Zürcher, & Burkart, 2017). Given the complexity of cognitive and social processes involved in the perception and production of speech, a graded view of automaticity could provide a useful framework for understanding alignment behavior (Bargh, 1994). That is, vocal alignment behavior could be recharacterized in terms of an automatic tendency to align and a secondary, goal-directed process that may serve to inhibit excess alignment (Bargh, 1994; Garnier, Lamalle, & Sato, 2013; Moors & De Houwer, 2006).

Bargh (1994) recommends a definition of automaticity in cognition that includes four components: intentionality, controllability, awareness, and efficiency. *Intentionality* and *controllability* are tightly linked constructs, the former referring to an individual's conscious desire to perform a behavior and the latter to an individual's ability to modify or inhibit a behavior once started. Because intentionality and controllability are so interrelated, it has been argued that they should be considered jointly as goal-dependent features of a process (Moors & De Houwer, 2006). *Awareness* refers to the conscious experience of and correct attribution of a process to its cause (Bargh, 1994). *Efficiency* pertains to whether or not effortful processing resources are required to initiate a process; an *efficient* process is one that occurs without cognitive or attentional resources. A main goal of the current dissertation was to test whether these components, particularly efficiency, can be used to describe or explain aspects of vocal alignment behavior.

It has been argued that an automatic entrainment process would be evolutionarily advantageous because it would confer all the communicative and social benefits without taking valuable attentional resources (van Baaren, Janssen, Chartrand, & Dijksterhuis, 2009; van Leeuwen, van Baaren, Martin, Dijksterhuis, & Bekkering, 2009). However, as van Baaren et al. (2009) point out, individuals do not constantly copy every behavior they observe. An obligatory, fully automated vocal alignment mechanism would be detrimental, as it would result in indiscriminate alignment (Brass et al., 2005; Spengler et al., 2010). If vocal alignment were unregulated, it would be impossible to observe distance or distinctiveness between individuals or speaking styles. A secondary mechanism may be in place to modulate the degree of alignment to others' speech (Garnier et al., 2013).

A candidate model for reconciling theories of vocal alignment and accounting for the body of evidence supporting each can be found in the domain of behavioral mimicry (Chartrand & Bargh, 1999; Chartrand, Cheng, Dalton, & Tesser, 2010; van Baaren et al., 2009). By such accounts, social stereotypes are primed automatically, and these stereotypic beliefs about a speaker can then in turn modulate the degree to which mimicry occurs. Automaticity can be understood here in terms of the definition described above: a process that is unconscious, efficient, and unintentional. However, a distinction can be made between the activation of stereotypes or beliefs and their *use* in guiding behavior. The authors suggest that behavioral imitation may be secondarily modulated by beliefs and stereotypes about the target of imitation. By this account, beliefs may serve as a "cognitive handbrake" on alignment (van Baaren et al., 2009; van Leeuwen et al., 2009). Alignment to speech occurs in parallel with alignment to non-speech behavior (Louwerse et al., 2012; Shockley et al., 2007) and proponents of a social mimicry approach implicitly assume speech is part of the suite of behavioral imitation (Chartrand & Lakin, 2013; Kurzius, 2015).

Relative Automaticity of Vocal Alignment

Reexamining vocal alignment in the context of relative automaticity (Schneider & Shiffrin, 1977), indirect evidence suggests that vocal alignment does not require an individual to be consciously aware of their behavior. For example, the ambient language or dialect environment can alter individuals' speech productions without directed attention (Delvaux & Soquet, 2007; Pardo, Gibbons, Suppes, & Krauss, 2012; Sancier & Fowler 1997). Evans and Iverson (2007) studied shifts in vowel production in students who moved to a university in a different dialect region. The authors recorded the students' productions of the same speech materials at regular intervals for the first two years of university. Acoustic analyses indicated that vowel formants shifted toward the community norms, but perceptual tests indicated that the students themselves failed to reliably detect the changes in their speech. As discussed earlier, vocal alignment has also been found in experimental paradigms that expose listeners to spoken words, suggesting that simply hearing the properties of an individual's voice is sufficient to induce alignment (Goldinger & Azuma, 2004; Fowler, Brown, Sabadini, & Weihing, 2003).

As a field, the goal-related aspects (e.g., intentionality and controllability) of vocal alignment have been widely studied and may be the most plausible explanation for vocal divergence (Coupland, 1984; Cutler, 2010; Giles et al., 1991; Giles & Ogay, 2007). Alignment has been demonstrated to be affected by the demands of a task, relative social status, and management of social distance (Aguilar et al., 2016; Babel, 2010; Bourhis & Giles, 1977; Coupland, 1984; Kim et al., 2011). In one example, participants in Babel's (2010) study of vocal alignment were either shown a photo (African American face or Caucasian face) or no photo prior to beginning a naming task. Females who rated the face as more attractive showed more alignment to vowel spectra, but males who rated the face as more attractive showed less alignment to vowels. The effect of explicit intent to imitate or to not imitate has also been studied. Explicit instructions to imitate a model talker yielded greater alignment on vowel spectra than no explicit instructions to imitate in an immediate shadowing task (Dufour & Nguyen, 2013). A separate study showed that, when participants were made explicitly aware of the vocal alignment phenomenon and were instructed to inhibit it, they were able to maintain their own characteristic voice properties (i.e., vocal pitch) or to diverge from the model talker's (Garnier et al., 2013).

Despite this evidence for control in vocal alignment, it is unclear from the current literature whether vocal alignment *requires* attentional or cognitive resources (i.e., is efficient; Bargh, 1994). One recent study by Abel and Babel (2017) suggests that cognitive resources are indeed required to initiate vocal alignment. This study examined vocal alignment in a dyadic setting where an instruction giver directed an instruction receiver in the construction of several Lego models of varying complexity. Perceptual assessments of alignment indicated that participants in the hardest condition showed the least alignment between task partners and those in the easiest condition showed the greatest alignment. These results suggest that when the task was harder and presumably required greater cognitive resources, speakers engaged in less vocal alignment. However, other recent work suggests that vocal alignment may actually *increase* as task difficulty increases (Solanki, Vinciarelli, Stuart-Smith, & Smith, 2016).

Solanki and colleagues (2016) examined vocal alignment in dyads using the UK Diapix task. The Diapix task consists of 12 paired images that differ on approximately six features. Participants must verbally collaborate to find the differences without being able to see their task partner or the partner's version of the image. Task difficulty was defined in terms of time to complete the task (i.e., "easy" picture sets took the least time to complete; "hard" picture sets took longest to complete), and alignment was assessed by acoustic similarity between task partners. The authors found that in the more difficult picture sets, where presumably cognitive load was higher, vocal alignment was greater between task partners. These results contrast with Abel and Babel (2017), leading to the seemingly paradoxical interpretation that higher processing demands both increase and decrease alignment and that vocal alignment both requires and does not require processing resources.

One potential issue with both studies is that the difficulty of the task may have been confounded by differences among the task participants themselves. In both experiments, difficulty was defined in part based on how effectively one partner described the features of an image. More challenging task segments were made so because participants had to describe subtler differences (Solanki et al., 2016) or more complex instructions (Abel & Babel, 2017). Thus, these tasks may have not only measured the effect of cognitive load, but also the effect of interlocutors' communicative skill. Particularly in Abel and Babel (2017), the cognitive load may have disproportionately fallen on the instruction giver. A more compelling test of processing demands should constrain cognitive resources independently of alignment. This was a primary goal of the current dissertation.

Individual Differences in Vocal Alignment

It is certainly not a matter of debate that individuals differ from one another in a number of ways. With respect to individual differences that affect speech, biological, situational, and social factors influence the realization of linguistic structure (Fitch & Giedd, 1999; Pennebaker, Mehl, & Niderhoffer, 2003; Smith & Patterson, 2005). The differences across individuals in speech may in turn contribute to the large individual differences observed in vocal alignment (Pardo, 2013; Yu, Abrego-Collier, & Sonderegger, 2013). Researchers have only recently begun to systematically examine the influence of individual differences on alignment behavior (Kurzius, 2015; Lewandowski & Nygaard, under review; Pardo et al., 2013; Yu et al., 2013).

There are two broad categories of individual difference that may account for variation in the tendency to align: perceptual-motor differences in the speech system, and personality or dispositional differences. These types of differences reflect physical and individual constraints on alignment. For example, Lewandowski (2009) examined vocal alignment of native German speakers to native English speakers in a dyadic task. The German speakers were classified a priori as high or low on “language talent” for English based on a battery of perceptual and productive measures (Jilka, 2009). Talented speakers of English showed greater alignment as measured by amplitude (e.g., loudness) than did the less talented speakers, demonstrating that basic differences in perceptual and motor speech ability may contribute to the extent of alignment.

Another class of individual differences concerns chronic or dispositional traits. A recent study examined the relationship between rejection sensitivity (RS) and vocal alignment (Aguilar et al., 2016). RS, or the extent to which individuals seek approval from others, mediated perceived level of closeness between participants in a map task (e.g., Anderson et al., 1991; Pardo, 2006) and affected degree of vocal alignment (Aguilar et al., 2016). When high-RS individuals were paired with low-RS partners, the high-RS partners aligned significantly more than the low-RS partners, despite the finding that vocal alignment was not significantly different across partners in matched (high-RS with high-RS partner; low-RS with low-RS partner) pairs. In mismatched pairs, the high-RS partner also reported less feelings of closeness. The results suggest that certain individual differences in disposition are associated with both vocal alignment and its downstream consequences.

Yu et al. (2013) examined the relationship between alignment to voice onset time (VOT; duration of a consonant closure before a vowel) and a host of individual traits.

Their battery included the Automated Reading Span Task (R-SPAN; Unsworth, Heitz, Schrock, & Engle, 2005) for working memory, the Big Five personality inventory (John, Donahue, & Kentle, 1991), and the Autism Spectrum Quotient (AQ; Baron-Cohen et al., 2001). Of these, only the Openness to Experience subscale of the Big Five was associated with vocal alignment (Yu et al., 2013; but see Kurzius, 2015). Taken together, these studies provide evidence that dispositional characteristics might be associated with the extent of vocal alignment. However, the individual differences tested to date are far from exhaustive, and other traits, such as empathy, may also be associated with variation in alignment behavior.

Empathy and its precursors have been linked to prosocial behavior in humans (Eisenberg & Miller, 1987) and in non-human primates (Campbell & de Waal, 2011; de Waal, 2008). For example, van Baaren, Holland, Kawakami, and van Knippenberg (2004) found that human participants whose postures were imitated by a confederate exhibited more helping behaviors (e.g., picking up dropped pens, donating money to charity) than participants who were not imitated. Individuals who express greater empathy and its associated cognitive precursors (e.g., perspective-taking) tend to demonstrate more behavioral imitation, making it potentially relevant for speech alignment as well (Chartrand & Bargh, 1999; Kühn et al., 2010; Meltzoff & Decety, 2003; Seyfarth & Cheney, 2013).

In the domain of speech, an effective listener seeks to understand a speaker's intentions by unpacking the propositional and emotional content of the speaker's utterances (Nygaard & Lunders, 2002; Nygaard & Queen, 2008). It follows that individual differences in the listener's ability to empathize with the speaker may relate to the manner in which individuals process the indexical and affective components of the speech signal (i.e., de Guzman, Bird, Banissy, & Catmur, 2016) and may have consequences for vocal alignment behavior. Indirect support for the role of empathy in vocal alignment comes from Neuman and Strack (1998). The authors presented participants with slightly happy, slightly sad, or affectively neutral speech. Although the emotional tone of voice was subtle enough that participants did not notice that emotional prosody was being manipulated, the shadowers nonetheless adopted characteristics of the emotional prosody and reported mood states in line with the prosody of the shadowed speech. This study did not examine individual differences in empathy, nor empathy per se. Rather, the authors examined emotional contagion, a related construct. However, their study suggests a role for empathic thinking or feeling in vocal alignment.

Overview of the Dissertation

The primary aim of this dissertation was to investigate the relative automaticity of vocal alignment and examine individual differences that may potentially be associated with the propensity to align. By introducing a cognitive load in a minimally social context, the following experiments evaluated the proposal that vocal alignment is an efficient process modulated by inhibitory, goal-related components. Two experiments assessed (1) the extent to which vocal alignment can be considered an efficient process (e.g., process that requires few cognitive resources; Bargh, 1994; Moors & DeHouwer, 2006), (2) whether goal-directed aspects of vocal alignment require cognitive resources to facilitate or inhibit vocal alignment, and (3) how individual differences in speech production and empathy are associated with alignment.

Experiment 1 directly tested the efficiency of vocal alignment by using a dual-task procedure to reduce the availability of cognitive or attentional resources. Experiment 2 investigated whether social information about a specific individual (i.e., information not gained through abstract or group stereotypes) would influence vocal alignment behavior. Experiment 2 evaluated whether using social information is a goal-dependent process that requires cognitive resources to influence the degree of vocal alignment. Within each experiment, assessments of individual differences in baseline variation in speech production are considered, and finally, the data from Experiments 1 and 2 were pooled to examine dispositional differences in empathy.

Experiment 1

Dual-task paradigms have been used to successfully tax attentional resources and disambiguate hypotheses regarding automaticity in a variety of psychophysical domains (Catmur, 2016; McGuire, Gillath, & Vitevitch, 2016; Saucedo-Marques, Ceux, & Wenderoth, 2011; Turk et al., 2013). For example, in an experiment examining behavioral mimicry, van Leeuwen et al. (2009) asked participants to tap their finger in response to a cue presented on an on-screen, computer-animated hand. While completing the tapping task, participants simultaneously performed a verbal 2-back task (Kirchner, 1958; Owen et al., 2005) in a high load condition or a 0-back task in a low load condition. More imitation (i.e., faster response to the cue) was observed in the high load condition than in the low load condition. The finding of increased motor imitation under cognitive load is consistent with the notion that behavioral imitation does not require attentional resources and furthermore that cognitive resources may be modulate imitation when processing demands are low (Garnier et al., 2013; Solanki et al., 2016). However, as noted earlier, when Abel and Babel (2017) manipulated processing demands of conversational dyads, they found decreased vocal alignment, suggesting that vocal alignment may not be an efficient process and is reduced when processing demands are high.

The primary goals of Experiment 1 were to (1) test whether or not vocal alignment requires cognitive resources and (2) examine whether social stereotypes affect vocal alignment behavior. By manipulating cognitive load, the possible mechanisms underlying vocal alignment can be evaluated. Depending on the extent to which vocal alignment is attention- or resource-demanding, three outcomes were posited for the current experiment. First, if vocal alignment is not efficient and therefore attentional resources are mandatory, then individuals may show reduced vocal alignment when fewer cognitive resources are available. This outcome would be consistent with Abel and Babel (2017). Second, if vocal alignment is an efficient process and therefore attentional resources are not mandatory, cognitive load should have no effect on alignment behavior, as limiting cognitive resources will not impact a process where no resources are necessary. This outcome would be consistent with accounts of vocal alignment that propose that properties of low-level cognitive or perceptual architecture may be sufficient to drive alignment (Fowler & Gallantucci, 2005; Pickering & Garrod, 2013; Shockley et al., 2007). Third, if vocal alignment is an efficient process modulated by a secondary, inhibitory mechanism, then adding load may increase alignment by reducing inhibition. This possibility would be consistent with work on non-verbal alignment behavior, and acknowledges that the mechanisms underlying alignment may be multi-faceted (Chartrand & Bargh, 1999; van Baaren et al., 2009; van Leeuwen et al., 2009).

The current experiment employed a dual-task structure similar to van Leeuwen and colleagues' paradigm (2009). While van Leeuwen et al. had participants complete a verbal n-back task and visuospatial motor task, participants in the current experiment completed a visuospatial n-back task (Owen et al., 2005) and a verbal naming task (e.g., Goldinger, 1998). Following Goldinger (1998), the naming task consisted of three kinds of trial blocks: a block in which participants say aloud words that they read on screen (baseline block), a block in which participants are perceptually exposed to the words and voices that they will hear in the final block (exposure block), and a block of trials in which participants say aloud words spoken by a model talker (shadowing block). Additionally, because participants were unlikely to have experience with the n-back task, a practice block was provided prior to beginning of the experimental tasks. A 1-back was chosen to achieve an appropriate level of difficulty for the participants (Gazzaniga, Ivry, & Mangun, 2009). Pilot subjects (n=3) all spontaneously reported that the task was extremely challenging for them.

Experiment 1 recruited model talkers from native and non-native language backgrounds, specifically, native speakers of English and native speakers of Mexico City Spanish. Non-native accents are both acoustically and socially distinct from standard dialects of a language (Giles & Ogay, 2007; Gluszek, & Dovidio, 2010). Social stereotypes associated with Spanish-accented individuals tend to be negative (McKirnan & Hamayan, 1984; Park-Taylor et al., 2008; Ryan, 1983; Ryan, Carranza, & Moffie, 1977), and isolated words are sufficient for listeners to make status and sociality judgments (Lewandowski & Nygaard, under review). It has been demonstrated in the literature that individuals align more to less intelligible and more distinctive speech (e.g. Babel et al., 2014; Kim, 2011; Walker & Campbell-Kibler, 2015), but also more to individuals with positive social attributions, (e.g. Babel, 2012; Babel & Bulatov, 2012; Bourhis & Giles, 1977; Gregory & Webster, 1996). Thus, Spanish-accented speech may provide an informative test case for social stereotype effects on vocal alignment. Specifically, vocal alignment may differ toward Spanish-accented and native-accented model talkers and cognitive load may differentially affect alignment to individuals who are perceived positively in comparison to individuals who are perceived less positively.

Vocal alignment was assessed by (1) independent raters' assessments of similarity between model talker and shadower utterances and (2) acoustic profiles of model talker and shadower utterances. This type of study design permits comparison across objective and perceptual measures of alignment. While it must be the case that listeners are holistically determining alignment based on acoustic parameters, individuals may be aligning and diverging on multiple dimensions simultaneously. To arrive at a more complete understanding of vocal alignment, it is necessary to be able to predict the specific dimensions along which alignment will occur. The current experiment examined alignment to duration (a proxy for speaking rate), F0 (pitch), and vowel spectra (spectral properties of vowels). These dimensions were chosen based on the likely parameters of alignment. Duration and F0 could be considered supra-segmental or even non-linguistic parameters of the speech signal, as individuals can adjust the duration and pitch of non-speech utterances as well, whereas vowel spectra properties index whether shadowers are aligning to the specific realization of linguistic structure.

A secondary goal of Experiment 1 was to examine the role individual differences in alignment behavior. As discussed in Chapter I (General Introduction), individual differences in speech production abilities and personality characteristics may affect intrinsic propensity to vocally align. Individuals differ in the habitual variability of their speech (Bradlow, Toretta, & Pisoni, 1996). Individuals with more baseline flexibility or variation in their speech may be better able to achieve target speech during vocal alignment because they have greater range in speech production (Babel, 2010; Walker & Campbell-Kibler, 2015). The current dissertation further built on previous work by exploring the association with individual differences in empathy. Empathic individuals are more sensitive to social cues than less empathic individuals, so individuals who score higher on empathy may be more likely to vocally align (Yu et al., 2013). If individual differences in baseline variation and empathy affect tendency to engage in speech alignment, these factors should affect vocal alignment regardless of cognitive load.

Methods

Participants

Independent participant groups were recruited from the Emory University community to complete the (1) naming task and (2) provide ratings of alignment. All participants were 18-25 years of age, native speakers of American English, and had no self-reported history of hearing or speech disorders. Specific exceptions are noted below. Participants received partial credit towards a psychology course requirement.

Shadowers.

Fifty-one shadowers (17 male, 33 female) participated in the naming task. Shadowers were randomly assigned to the cognitive load (CL) or the no load (i.e., no distractor; ND) condition. Eleven participants were excluded due to computer technical error (4), history of speech disorders (4), experimenter error (2), or failure to complete the experimental tasks (1). Data from the remaining 40 shadowers were analyzed. Of these, 20 shadowers (6 male, 14 female) were assigned to the CL condition and the remainder, to the ND condition (10 male, 10 female).

AXB raters.

Two hundred and twenty-four independent raters (five or six raters per shadower) provided perceptual assessments of vocal alignment. Data from 21 raters were excluded because of computer technical error (10), history of hearing or speech disorders (9), language background (e.g., non-native English speaker, 1), or failure to complete the experimental task (e.g., fell asleep during the task, 1). The remaining 203 raters were all native speakers of English, 52 of whom were also speakers of another language. The following languages were represented within the bilingual AXB raters: 12 Spanish, 9 Chinese (3 Mandarin), 6 Korean, 2 Creole (unspecified types), 2 Farsi, 2 French, 2 Hindi, 2 Telugu, 2 Urdu, 2 Vietnamese, 1 Arabic, 1 Gujarati, 1 Italian, 1 Nepali, 1 Norwegian, 1 Russian, 1 Serbian, 1 Tagalog, 1 Tamil, 1 Turkish, and 1 Twi. None of the raters had participated as a shadower.

Materials

Word stimuli.

Seventy-two words were selected from the Emory University Speech and Language Perception Lab's database of foreign-accented and native English (see Sidaras et al., 2009) on the basis of frequency and neighborhood density. Half the items were low frequency ($\bar{x}$=12.22; Kučera and Francis, 1967) CVC (consonant-vowel-consonant) words with many ($\bar{x}$=282.22) high frequency neighbors, and the remainder were high frequency ($\bar{x}$=309.69) CVC words with few low frequency neighbors ($\bar{x}$=38.32; Luce & Pisoni, 1998). Fifteen monophthongal and diphthongal vowels were represented unequally in the word set. The following list describes the word set's vowels and, to facilitate comprehension, includes a sample word for each vowel category: 9 /ɪ/ (f*i*g), 8 /æ/ (ch*a*t), 7 /i/ (b*ea*n), 6 /eɪ/ (g*a*ve), 6 /u/ (d*oo*m), 6 /ʌ/ (h*u*m), 5 /a/ (b*a*lm), 5 /ɔ/ (th*ou*ght), 5 /ɝ/ (c*ur*ve), 3 /ɛ/ (p*e*t), 2 /aʊ/ (m*ou*th), 2 /oʊ/ (g*oa*t), 2 /ɔɪ/ (j*oi*n), 2 /aɪ/ (l*i*ce), 2 /ʊ/ (p*u*t). It should be noted that phonetic vowel category is not well-represented by Standard English orthography. Appendix A contains the full word list.

Shadower utterances were recorded in a sound-attenuated room using a head-mounted Sennheiser HMD-280 (30Ω) microphone onto a Dell computer and automatically segmented by trial using Eprime 2.0 (Psychology Software Tools, Inc., 2002). The sound files were digitized at 22.050 kHz and amplitude normalized in SoundStudio. During sound processing, research assistants blind to condition marked utterances on which shadowers made a naming error (e.g., "goad" instead of "goat"). Naming errors accounted for a small percentage (3.72%) of all responses.

Model talkers.

Four talkers from the Speech and Language Perception Lab's speech database served as target speakers for the current experiment. English Male (EM) and English Female (EF) were the native targets (L1=Standard American English), and Spanish Male (SM) and Spanish Female (SF) were the non-native targets (L1=Mexico City Spanish). SM arrived in the United States (US) at age 22, began learning English at age 20, and had lived in the US for 7 years at time of recording. SF arrived at age 34, studied English since age 2, and had lived in the US for 3 years at the time of recording. Independent ratings confirmed that the two native speakers were more intelligible, less accented, and perceived as higher status than the two Spanish speakers (see Table 1; ratings procedure described in Lewandowski & Nygaard, under review). Model talker utterances were used as target stimuli in the naming task and as AXB comparison stimuli.

Table 1. *Intelligibility, accentedness, and attitudes ratings of the model talker stimuli. Intelligibility is measured as transcription accuracy. Mean accentedness scores based on a 1-7 Likert type rating, where higher numbers indicate greater degree of foreign accentedness. Sociality and Status composite Likert ratings of the 21 attributes collapsed into categories based on the results of factor analysis (Reproduced with permission from the authors; Lewandowski & Nygaard, under review).*

Model Talker	Baseline Intelligibility (% correct)	Accentedness Rating (1-7 scale)	Sociality Rating (1-7 scale)	Status Rating (1-7 scale)
English Female	94	1.89	4.43	5.17
English Male	93	1.92	4.23	4.10
Spanish Female	66	5.63	4.69	4.08
Spanish Male	54	4.77	3.97	3.84

Image stimuli.

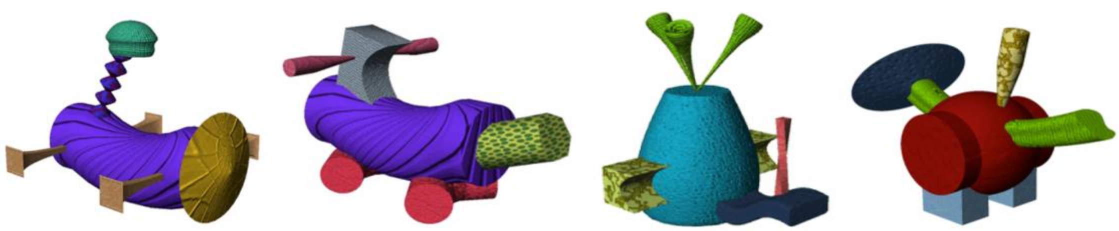

Figure 1. Sample Fribble images used in the visual 1-back task. Stimulus images courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University. Reproduced under a Creative Commons 3.0 License.

Fribbles (Williams, 1998) were chosen as 1-back image stimuli because they are novel, computer-generated creatures that discouraged participants from covertly labelling and rehearsing the image labels (e.g., Figure 1). Fribbles are grouped into three families based on their body style and color. While it is possible to generate labels for individual Fribble parts (e.g., “purple”, “samurai hat”), labelling the Fribble would be difficult under the time constraints imposed by the experimental task. Eighty unique Fribble images (27 purple family, 27 blue family, 26 red family) were resized to 400 x 400 pixels at 72 ppi resolution in RGB-8 color space and PNG-compressed. Fribbles always appeared against a white background.

Empathy Quotient.

To measure empathy, the Baron-Cohen and Wheelwright (2004) Empathy Quotient (EQ) was administered to shadowers. Unlike many other measures of empathy, this scale specifically targets empathic thinking and feeling, rather than general emotional reactivity (e.g., Questionnaire Measure of Emotional Empathy, Mehrabian & Epstein, 1972; Mehrabian, Young, & Sato, 1988) or social skill (e.g., Empathy Scale, Hogan, 1969; Davis, 1994). There are no subscales on the EQ, as the cognitive and affective components of empathy are highly correlated and difficult to disentangle (Baron-Cohen & Wheelwright, 2004; but see Davis, 1980 for an attempt).

The EQ contains 40 empathy statements (e.g., *I find it easy to put myself in somebody else's shoes, People sometimes tell me that I have gone too far with teasing*) and 20 filler statements (e.g., *I think that good manners are the most important thing a parent can teach their child)*. Participants respond on a 4-point Likert-type scale from "strongly agree" to "strongly disagree." Of the 40 empathy statements, roughly half require responses of "slightly/strongly disagree" in order to count toward the empathy score while the remainder require responses of "slightly/strongly agree." The maximum possible score is 80 and the minimum is 0, with 40 representing the approximate midpoint. Baron-Cohen and Wheelwright (2004) report that males score lower on average than females by about 5.4 points. Although this measure was collected for Experiment 1, an analysis of the relationship of empathy and vocal alignment is not presented in this chapter. To improve statistical power for the detection of individual differences in disposition, a pooled analysis of empathy was performed on the combined data from Experiments 1 & 2 and will be reported in the next chapter.

Procedure

Naming task.

The general naming task procedure will be described, followed by the differences between the CL and ND conditions (see Figure 2 for schematic overview). Participants completed the naming task in a single testing session. First, participants completed a block of 20 1-back practice trials regardless of whether they were assigned to the CL or ND conditions. This was done to equate task length across load conditions. Participants received corrective feedback throughout practice.

The first block of the experimental task consisted of baseline trials in which participants read aloud 72 words as they were presented one at a time on a computer screen. Following the baseline block, participants engaged in a perceptual exposure block. On each trial, participants listened to a word spoken by one of the four model talkers and selected the target word from an onscreen table of nine choices. The closed-set choice task was included to ensure participants attended to the words. The final block was the shadowing block, in which participants listened to the 72 words spoken by model talkers and said them aloud.

Participants listened to the model talkers produce a unique subset of 18 words (e.g., "balm" spoken only by EF) during the perceptual exposure and shadowing blocks. Model-word pairings were counterbalanced across shadowers with a Latin-square design. Stimulus presentation in all blocks was fully randomized (i.e., not grouped by speaker or word). Task instructions specifically excluded words like "imitate" or "repeat" to avoid biasing participants toward vocal alignment. After completing the naming task, a research assistant administered the Empathy Quotient questionnaire (EQ), and then debriefed the participant.

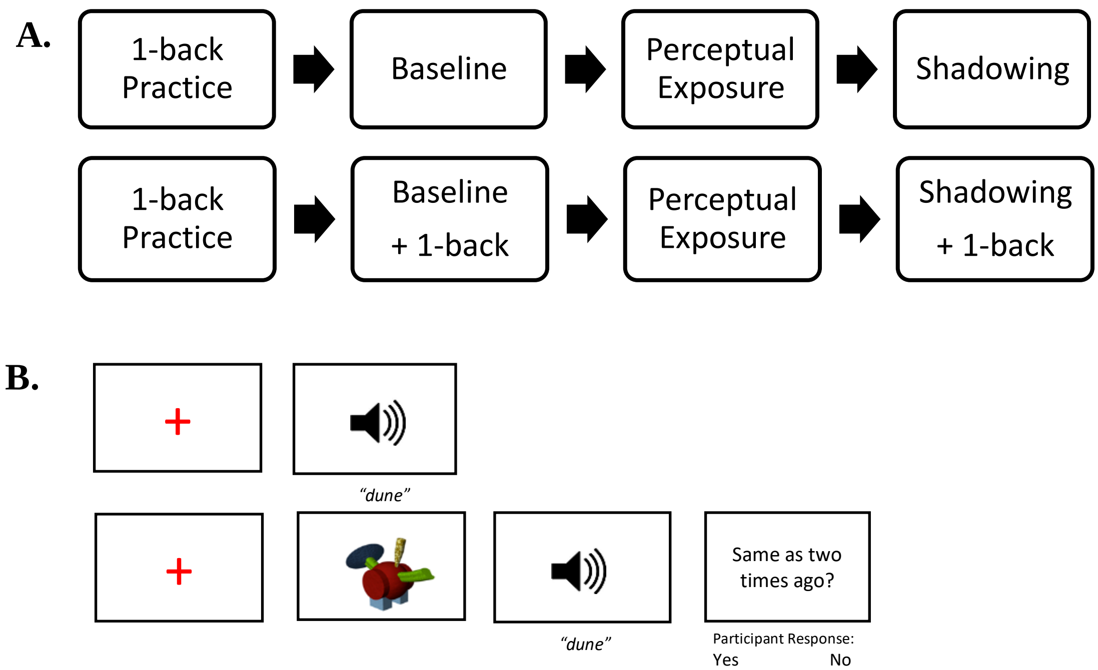

Figure 2. The experiment and trial structure for the naming task. (A) The block progression of the experiment is identical across ND and CL conditions, with the key difference being the dual-task nature of the baseline and shadowing blocks. (B) depicts the typical trial structure for the shadowing block. Fribble image courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University.

The crucial difference across conditions was the presence or absence of cognitive load. The CL group completed the baseline and shadowing blocks with a concurrent 1-back task while the ND group does not. On each baseline and shadowing trial, participants in the CL condition view a fixation cross (500 ms), then a Fribble (500 ms).

Participants were then presented with either a printed word (baseline block) or spoken token (shadowing block) and given a 2000 ms spoken response window. Following the spoken response, participants had unlimited time to make their yes/no decision for the 1-back. Because participants could not make their response until after they completed a naming trial and therefore were required to update the images serially in working memory, a cognitive load was present during naming.

AXB perceptual rating task.

Shadowers may align or diverge simultaneously on many acoustic dimensions and therefore, an AXB rating task was conducted to obtain a global assessment of vocal alignment (Goldinger, 1998; Miller, Sanchez, & Rosenblum, 2013; Pardo et al., 2013b). Utterances from the baseline and shadowing blocks of the naming task were segmented into word-length sound files by research assistants blind to condition and then the segmented utterances were used as stimuli in the AXB task. On each trial of the AXB task, raters judged whether the shadower's shadowed utterance (A) or the shadower's baseline utterance (B) sounded more similar to the model's talker target (X). If vocal alignment has occurred, raters should report a greater similarity between the shadower's shadowed production and the model talker target (AX) than between the shadower's baseline utterance and the model talker target (BX) at rates that are reliably greater than chance (.50). The order of A and B were counterbalanced across rating trials.

Results

The data were analyzed using mixed effects regression modeling (MEM; Baayen, Davidson, Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013) in `R` (R development core team, 2013) using the `afex` (Singmann et al., 2016) and `lme4` (Bates, Maechler, Bolker, & Walker, 2014) packages. This technique allows effects of interest (i.e., fixed effects) to be considered alongside multiple sources of random variation (i.e., random effects) in the data. For analysis of acoustic measures of alignment, general linear mixed effects models (GLMM; Baayen & Milin, 2015) were used. Where appropriate, logistic mixed effects modeling was used to analyze data bounded by 0 and 1 (AXB measure) because other modeling techniques can over- or under-estimate effects when the outcome measure nears a boundary (Jaeger, 2008).

All MEMs utilized the maximal random effects structure justified by the data, including random slopes and intercepts where appropriate[1] (Barr et al., 2013). The significance of fixed effect factors was justified both within a model with *t* (for GLMM) and *z* (for LMER) statistics, and with stepwise model comparisons using AIC and BIC fit statistics. This testing procedure ensures that a fixed effect not only has explanatory power within an environment, but also that the inclusion of the fixed effect provides an overall better fit to the data. The first step was always to generate a *control MEM*, or a MEM that contained only the random effects, then to add factors of interest to evaluate whether each fixed effect and interaction offered a better fit.

Categorical fixed effects (model talker, cognitive load) are treatment coded (-.5, .5) in all analyses. With this coding approach, β-weights can be interpreted as the

[1] Random effects (RE) structure for analyses of perceptual measures: (1+accent|shadower) + (1+AXB order|AXB rater) + (1|word). RE structure for analyses of acoustic measures: (1+accent|shadower) + (1|word). Words were nested within shadower, making the random slope (1+shadower|word) a potential alternative to the random intercept (1|word). However, a control MEM with the former RE structure failed to converge, indicating over-parameterization (Jaeger, 2009), and was consequently discarded.

estimate of the fixed effect versus the unweighted grand mean (Keppel & Wickens, 2004). Dummy coding (0, 1) permits analyses of simple main effects, but not true main effects. For this reason, treatment coding is also preferable to dummy coding for testing interaction effects (Keppel & Wickens, 2004; Kugler, Trail, Dziak, & Collins, 2012).

Cramér's V was used to assess multicollinearity between nominal variables (`vcd`; Meyer et al., 2016), inter-class correlations to assess association between nominal and numeric variables (`ICC`; Wolak, 2016), and kappa for association among numeric variables (`LanguageR`; Baayen, 2011). Multicollinearity was generally low to moderate (.003 < Cramér's V < .06; .00003 < ICC < 0.14; 1.14 < κ < 1.36). There were a few general exceptions. Shadower was perfectly associated with condition and AXB rater, and model talker was perfectly associated with accent (all Cramér's V=1). Other specific, relevant exceptions are noted in the sections that follow.

Preliminary Analyses

To assess whether the 1-back task effectively taxed participants' cognitive resources, naming RT and accuracy were examined across conditions. Response times (RT) equal to 0, which indicate an RT recording error, and RTs ±2 standard deviations away from a shadower's mean RT were excluded from the analysis (4.2% of the data). When entered into an MEM as a fixed effect, it was revealed that RT was marginally slower in the CL condition ($\bar{x}_{CL}$=1049.44) than in the ND condition $\bar{x}_{ND}$=(937.15 ms), β_{CL}=111.65, t=1.75, p=0.0885. Although condition was only marginal within the MEM, the MEM that included condition significantly outperformed the control MEM, $\chi^2(3)$=2555.80, p<.001, suggesting that condition accounted for additional variance in RT relative to random effects alone. Total number of naming errors did not significantly differ across conditions (ND=50; CL=76), $\chi^2(1)=.51$, $p=.474$. Within the CL condition, mean accuracy on the 1-back itself was .70, but was variable across subjects, SD=.17, range= .36-.90. In summary, the range of performance on the 1-back task within the CL group and the overall slower RTs suggests that the cognitive load task was effective. The lack of differences across conditions in naming errors indicated that shadowers in the CL condition attended to both tasks.

Cognitive Load and Vocal Alignment

Perceptual assessment.

Prior to analysis, AXB responses were recoded as "1" when the shadowed utterance was rated more similar to model talker utterance and "0" when the baseline was rated more similar to model talker utterance, resulting in a proportion of perceived vocal alignment. If significant vocal alignment occurred, then raters should judge shadowed utterances as more similar to the model utterances at levels significantly above chance (.50). The intercept of the control MEM was significant ($\beta= 0.24$, $z= 4.64$, $p<.001$), indicating that perceived alignment ($\bar{x}=.56$) reliably differed from chance. To test the effects of cognitive load on vocal alignment, the primary factors of interest were cognitive load, model talker's accent, and their interaction. However, due to the small number of model talkers, alignment to individual model talkers was also examined.

Although multi-collinearity in MEMs examining perceptual alignment was generally low, there were some exceptions. AXB rater (subject) was moderately associated with perceived vocal alignment (ICC = 0.13), suggesting that raters may have differed in their sensitivity to vocal alignment (Namy et al., 2002). Associations between fixed effects of interest and random effects may underestimate the relative contribution of either.

Figure 3 depicts the relationship among accent (3A), model talker (3B), cognitive load, and perceived vocal alignment. The main effect of cognitive load on perceived alignment was non-significant, β_{CL}=0.02, z=.22, p=.83, and did not improve MEM fit over the control MEM, $\chi^2(1)$=0.05, p=0.83. A marginal main effect of accent was observed, $\beta_{Spanish}$= 0.07, z=1.75, p=.08. Because categorical fixed effect factors were treatment coded, the β-estimate indicates that there was overall greater alignment to Spanish-accented model talkers. The MEM containing accent as a fixed effect marginally outperformed the control MEM, $\chi^2(1)$=2.95, p=0.086. However, the accent x cognitive load interaction term was neither significant β_{CLxSpn}=-0.04, z=-.50, p=0.615, nor did it improve model fit over an MEM that did not include the interaction term, $\chi^2(1)$=0.25, p=0.615.

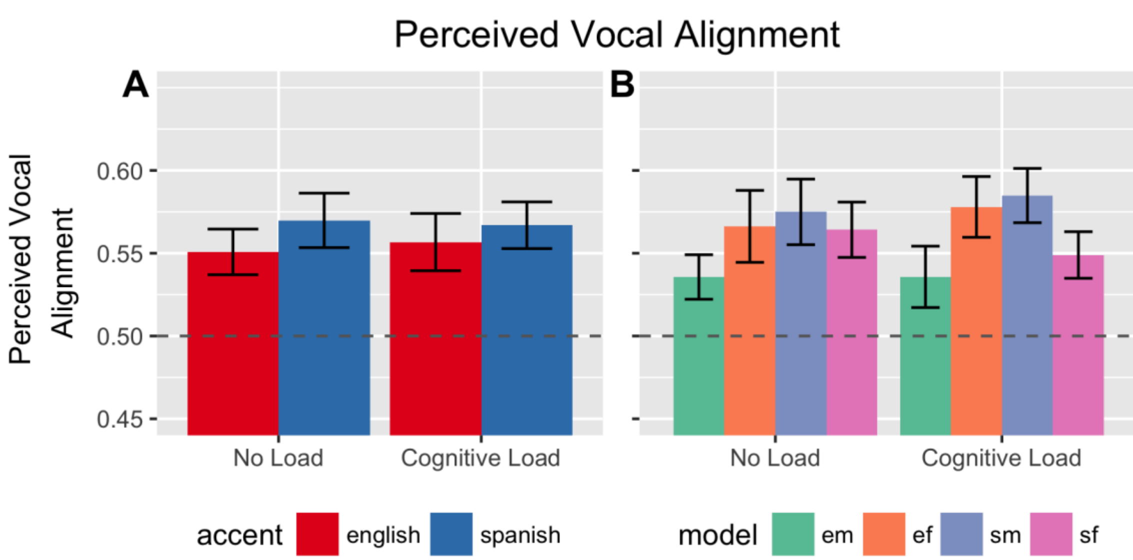

Figure 3. The effects of cognitive load and accent (left) and cognitive load and model talker (right) on perceived vocal alignment. The dashed line indicates the minimum threshold for either alignment or divergence. Unless otherwise specified, all error bars represent ±standard error of the mean (SEM).

When model talker was entered into the MEM in place of accent, a significant effect of model talker was observed, $\chi^2(3)=64.40$, $p<.001$. Alignment to all four model talkers exceeded chance. Inspection of the β-weights indicates that alignment to EF and SM was greater than the unweighted grand mean, $\beta_{EF}=0.10$, $z=2.01$, $p=.042$; $\beta_{SM}=0.18$, $z=3.73$, $p=0.0002$. Alignment to SF did not differ from the unweighted grand mean, $\beta_{SF}=-0.04$, $z=-0.77$, $p=0.441$. Alignment to EM was significantly lower than the grand mean $\beta_{EM}=-0.24$, $z=-5.01$, $p<0.001$. No interaction between model talker and cognitive load was observed, all $\beta< 0.09$, $z< 0.89$, $p>0.175$; $\chi^2(3)= 4.06$, $p=0.255$. The findings suggest that degree of alignment differed across model talkers, but not by load condition.

In order to ensure that the null effect of cognitive load on perceptual assessment of alignment was not due to an insufficient cognitive load, a follow-up analysis examining the relationship between 1-back performance and perceived alignment was carried out on the CL condition's data. There was a statistically reliable, positive association between 1-back accuracy and perceived alignment, $\beta_{Accuracy}=0.07$, $z=2.17$, $p=.030$. The addition of 1-back accuracy also improved MEM fit to the data, $\chi^2(1)=4.68$, $p=.030$. As shadowers' performance on the 1-back task increased, listeners perceived more vocal alignment. Higher accuracy on the 1-back task may reflect increased attention or cognitive effort directed to the task, which is consistent with the hypothesis that higher processing demands unmasked automatic processes associated with vocal alignment. Because shadowers' higher accuracy was associated with increased perception of alignment, the results suggest that cognitive load task was at least somewhat effective.

Acoustic assessment of vocal alignment.

Targeting specific measures provides insight as to which properties of the speech signal shadowers will align. Fundamental frequency (f0; pitch), utterance duration (speaking rate), and vowel spectra (F1 & F2)[2] were measured in Praat using custom scripts (Boersma & Weenink, 2014). F1 and F2 were calculated from the midpoint of the vowel and a mean F0 was calculated over the entire utterance. Both were collected using an automated LPC analysis and were normalized to allow comparison across talkers while minimizing variance not due to idiolect or dialect. A z-normalization method was used for pitch (F0) and Labov normalization was used for vowel spectra (F1 & F2; Labov, Ash, & Boberg, 2006; Thomas & Kendall, 2015). Inaccuracies (measures >2 SD away from an individual's mean F1 or F2) were replaced with manual measurements.

To index vocal alignment, difference in distance (DID) scores were calculated (Lewandowski & Nygaard, under review; Pardo et al., 2013). DID is computed according to equation (1), where B indicates the shadower's baseline value, M indicates the model talker's value, and S indicates the shadower's shadowed value:

$$DID = |M - B| - |M - S| \tag{1}$$

If a shadower exhibits alignment along a particular acoustic dimension, $|M - S|$ should be smaller than $|M - B|$, and the overall DID score should be positive. A negative DID

[2] *Vowel formants* are spectral peaks in the resonance of the vocal tract and serve as cues to vowel and speaker identity (Titze et al., 2015). Formant frequencies are canonically measured in Hertz (Hz). "F1" indicates the lowest Hz resonant frequency, "F2", the next highest, etc.

score would indicate divergence. The DID's magnitude reflects how much the shadower shifted toward or away from the model talker's production.

Duration.

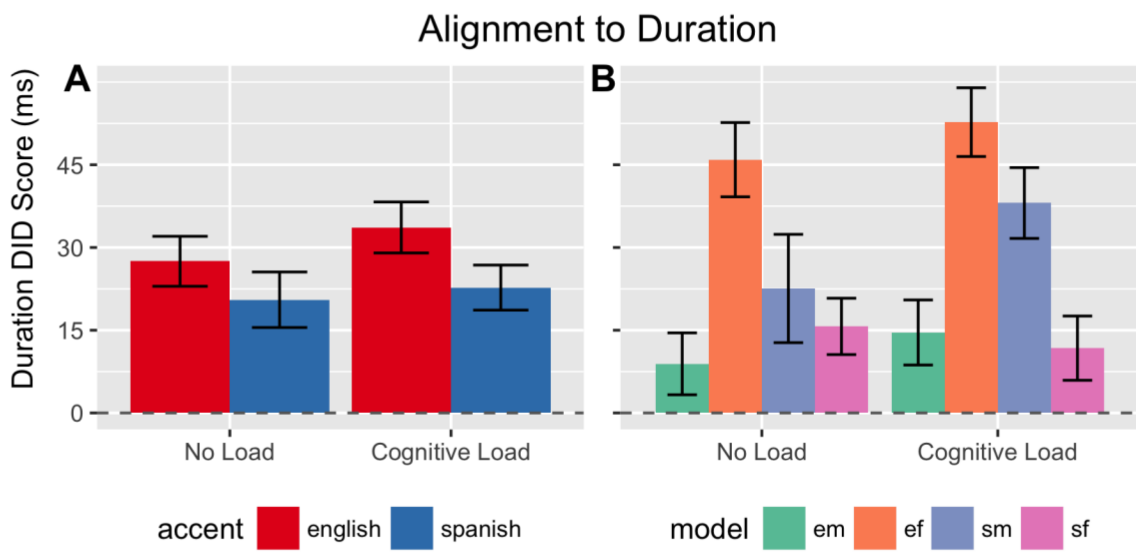

Figure 4. The relationship between cognitive load and accent (left) and cognitive load and model talker (right) in alignment to duration.

The associations among duration DID, cognitive load, accent and model talker are presented in Figure 4. The numeric trend suggests that there is greater alignment under cognitive load, and the effect of cognitive load on duration DID is statistically reliable, β_{CL}=47.85, t=2.08, p=0.039; however, adding condition as a fixed effect factor did not significantly improve fit over the control MEM ($\chi^2(1)$=0.43, p=0.512). There was no significant effect of accent on duration alignment, $\beta_{Spanish}$=-5.45, t =-1.40, p =0.171. Accent did not improve the MEM's fit to the data, $\chi^2(3)$=1.90, p= 0.168. Although the effect was not significant, the direction of the β-estimate indicates that there was generally less alignment to the duration of the Spanish-accented talkers than to the native model talkers. The condition x accent interaction did not reach statistical significance, β_{CLxSpn}=-5.75, t=-0.739, p=0.464; $\chi^2(1)=0.54$, $p=0.461$.

When model talker identity replaced accent in the MEM, model talker significantly improved MEM fit over the control MEM, χ^2 (3)=64.42, $p<.001$. The significant influence of model talker indicates that patterns of alignment to duration were not equivalent across target speakers. Alignment to EF's speaking rate exceeded the grand mean, β_{EF}=38.40, t=6.299, $p<.001$, and alignment to SM was marginally greater than the mean, β_{SM}=11.90, t=1.95, p=0.0523. Shadowers aligned least to EM's and SF's speaking rates, β_{EM}=-27.40, t=-4.51, $p<.001$ β_{SF}=-23, t =-3.75, $p<0.001$. The model talker x cognitive load interaction did not significantly impact duration alignment (all βs<9.95, t< 0.83, p>0.407). The model talker x condition interaction did not improve the MEM's fit to the data, $\chi^2(3)=1.19$, p=0.755.

Fundamental frequency.

Relationships among z-score normalized F0 DID and the variables of interest are shown in Figure 5. For z-normalized F0 DID, there was a significant main effect of load condition, β_{CL}=0.07, t =2.31, p=0.026. The β-estimate suggests that alignment in the CL condition was greater than the grand mean. The MEM containing load condition differed significantly from the control MEM, $\chi^2(1)=4.95$, p=0.0261. No main effect of accent was observed, $\beta_{Spanish}$= 0.022, t =0.88, p =0.379; $\chi^2(1)=.77$, p=.38, and the load x accent interaction was not significant, β_{CLxSpn}=0.044, t =0.88, p =0.378; $\chi^2(1)=0.78$, p=0.378. The analysis was repeated with model talker instead of accent. When model talker was included in the MEM, the main effect of cognitive load remained, but no significant effect of model talker was found (βs<0.03, *t*<0.79, *p*>0.430; $\chi^2(3)=.899$, *p*=.826). The analysis did not show a significant model talker x cognitive load interaction, all βs<.001, *t*< 0.15, *p* > 0.263; $\chi^2(3)=1.27$, *p*=0.737.

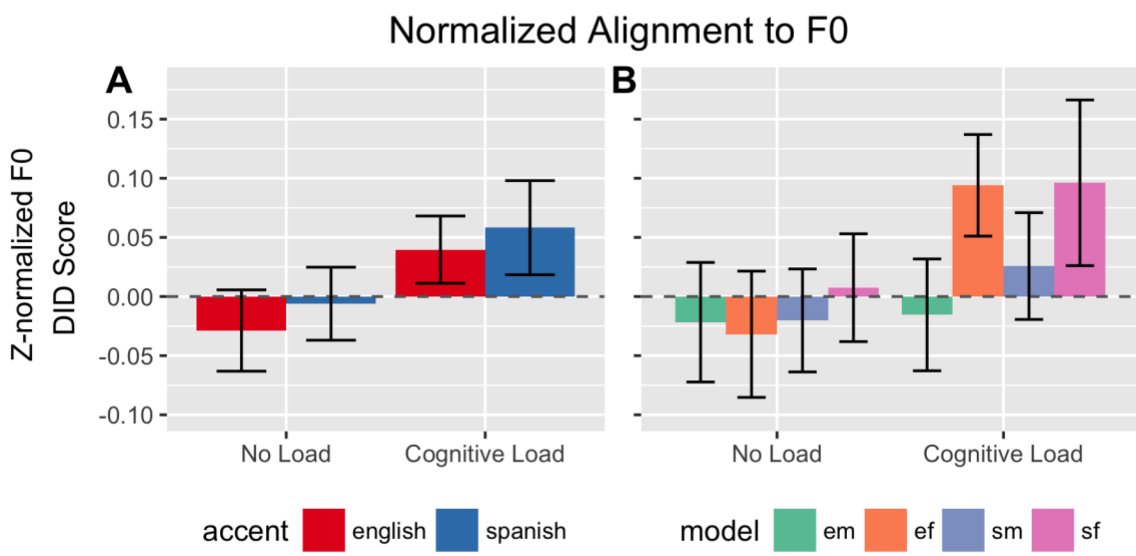

Figure 5. The relationship between accent and cognitive load (left) and model talker and cognitive load (right) and F0 alignment.

Vowel spectra.

The final acoustic parameter of interest was alignment to Labov-normalized vowel spectra. Figure 6 depicts the relationships among vowel DID, cognitive load, accent (6A) and model talker (6B). The visual trends in 6A suggest that shadowers align to native model talker vowel spectra and diverge from Spanish-accented vowel spectra in the ND condition, and this pattern is attenuated in the CL condition. Under cognitive load, shadowers show modest alignment to SM, but neither align nor diverge from the other three talkers. The main effect of cognitive load on alignment to vowel spectra was not significant, $\beta_{CL}=-11.57$, *t* =-.707, *p*=.484. Including cognitive load as a fixed effect did not improve fit relative to the control MEM, χ^2 (1)=.438, *p*=.508.

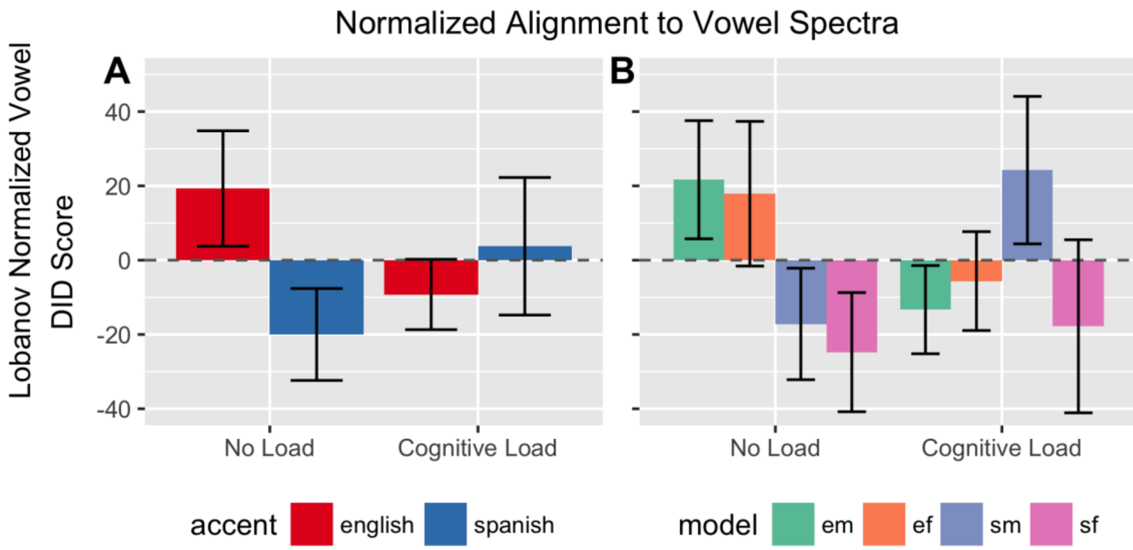

Figure 6. Interaction between accent and cognitive load (left) and model talker and cognitive load (right).

Although the test of cognitive load was inconclusive, the effect of cognitive load may have differed across accents, which could implicate social stereotypes in guiding alignment behavior. Although no significant effect of accent on vowel spectra alignment was observed, $\beta_{Spanish}$=-12.26, t=-1.12, p=0.264; $\chi^2(1)$=1.18, p=0.277, the load x accent interaction was significant, $\beta_{CLxSpanish}$=49.28, t=2.25, p=.026, and improved MEM fit ($\chi^2(1)$=4.98, p=0.026). The results indicate that alignment to vowel spectra for talkers of varying language backgrounds was affected by cognitive load. It is also apparent from visual inspection that there were differences not only by accent group but also by individual model talker. There was no main effect of model talker, all β<2.53, t<1.62, p>0.11, but the model talker x cognitive load interaction marginally improved MEM fit, $\chi^2(3)$=6.76, p=.08. Follow-up comparisons indicated that the interaction was primarily driven by SM, whose vowel spectra received significantly greater alignment in the CL condition than the other three talkers', β_{SM}=87.60, t =2.42, p=.016. No other pairwise comparisons were statistically reliable, all β<60.13, p>.10.

In addition to potential differences across *individual* model talkers, differences in alignment to vowel spectra may emerge due to *linguistic* factors. The spectral properties of vowel categories may differentially lend themselves to alignment (Babel, 2012). Because there were 15 vowels represented in the word set, a by-vowel analysis would have led to a substantial multiple comparisons problem (e.g., 5 Type I errors guaranteed at α=.05) or a substantial reduction in power to detect real effects. The current analysis grouped vowels together by articulation characteristics, tongue position, and lip roundness (close front, 16 words; close back, 8 words; open front, 21 words; open back, 11 words; diphthong, 16 words), reducing the number of comparisons from 15 to 5. It was also of interest to establish whether there were differences in alignment to the spectra of vowel categories across accent group. There are well-characterized differences between English and Spanish vowel inventories (Bradlow, 1995), which may give rise to such differences. The relationship between vowel DID and vowel category is shown in Figure 7.

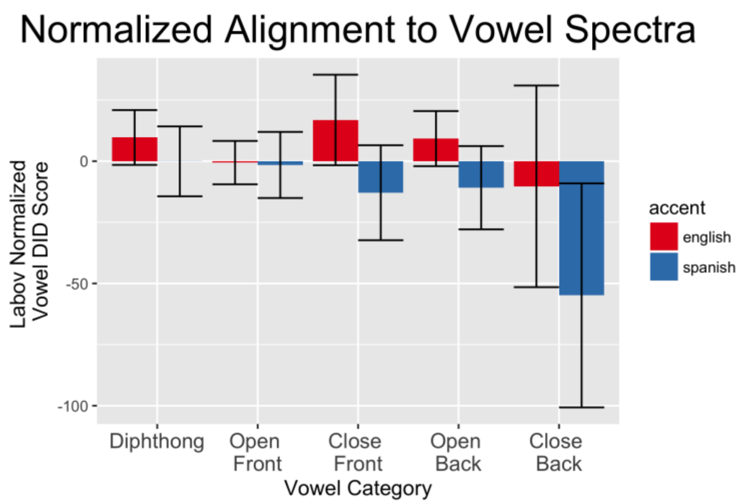

Figure 7. Alignment to vowel spectra as a function of vowel category and accent. There is a high degree of variation within each category, making it difficult to see patterns across vowel categories.

When vowels were collapsed across phonological category, no significant effect of vowel category on vowel DID emerged, $\beta<25.22$, $t<1.15$, $p>0.255$; $\chi^2(4)=2.47$, $p=0.65$). The vowel category x accent interaction was not justified by inspection of β-weights (all βs<39.15, $t<2.41$, $p>0.10$) or by MEM comparisons ($\chi^2(4)=7.7$, $p=0.11$). High variability within and between vowel categories make it difficult to draw firm conclusions, but the data suggest that the phonological content, as assessed by macro phonological features, may not have specifically influenced patterns of convergence or divergence from vowel spectra, at least for these speakers and vowel categories.

Individual Differences

Researchers have suggested that shadowers may be more likely to vocally align if the target is within their habitual production range (Babel, 2010; Babel, 2012; Kim, 2012; Lewandowski & Nygaard, under review; Walker & Campbell-Kibler, 2015). This proposal has received some support for cross-dialect shadowing, but less support for cross-accent shadowing. For example, Walker and Campbell-Kibler (2015) examined convergence on vowel spectra of native speakers of English from four dialect backgrounds: New Zealand, Australia, U.S. Inland North and U.S. Midland. The authors found that shadowers generally aligned to the spectra of vowels from the linguistically more distant dialect. However, Kim et al. (2011) found that vocal alignment in a dyadic task was greatest when interlocutors shared a common dialect. Although the conclusions are disparate, the literature suggests that vocal alignment may be greatest when the target is farther from the norm but still within the talker's production repertoire. However, past work has defined 'productive repertoire' very broadly.

Following Lewandowski and Nygaard (under review), 'variability' in production (or articulatory flexibility) was operationalized as standard deviation (SD) for duration and F0 and as vowel dispersion for vowel spectra. These measures were calculated on a by-participant basis from baseline utterances[3]; additional details on vowel dispersion will be provided in the relevant section below. Mutlicollinearity for this set of analyses was moderate (4.79 < κ < 10.30; .0001< ICC < 0.05), but there were some exceptions. Shadower was perfectly associated with all measures of variability (ICC = 1), and vowel dispersion was strongly associated with word (ICC=0.45).

Variation in duration and fundamental frequency.

Relationships between SD and acoustic DID for duration and F0 are shown in Figure 8. From a visual inspection, greater variability in baseline duration seems to be associated with less alignment to speaking rate while the opposite appears to be the case for pitch. When SD is entered into an MEM predicting the relevant DID score, neither trend is statistically reliable (β_{DurSD}=-0.34, t=-1.46, p=0.151; β_{F0SD} =0.03, t=-0.89, p=0.38). Baseline variability also did not improve fit for either MEM (Duration SD:

[3] The reader should note that the DID scores were also calculated, in part, from the baseline utterances. This may raise concerns about statistical independence (i.e., circular reasoning) of the independent variable and dependent measure; however, although the data source overlaps, the *type* of measurement is independent. This is similar to the way in which a mean and a standard deviation may be computed from the same data, but are independent constructs.

$\chi^2(1)=2.08$, $p=0.15$; F0 SD: $\chi^2(1)=0.77$, $p=0.381$). These findings indicate that variability in habitual speaking rate and pitch were not associated with vocal alignment.

Variation in vowel spectra.

Although standard deviations were a sufficient proxy for variability on duration and F0, vowel spectra are multi-dimensional. Vowel formant structure is characterized by

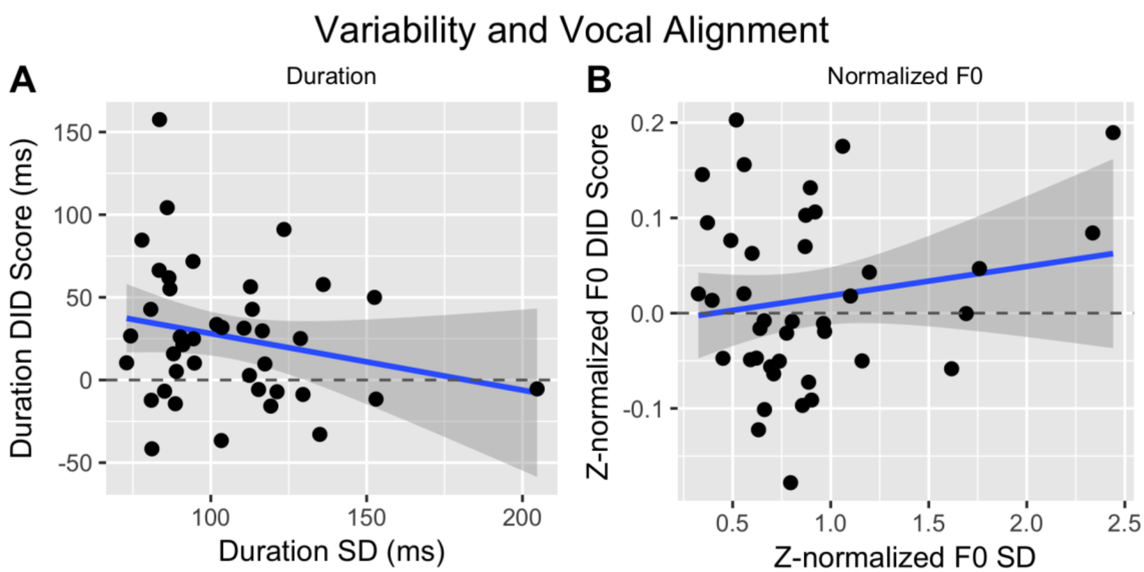

Figure 8. The relationship between variability in and alignment to duration (left) and the relationship between variability in and alignment to F0 (right). Points represent individual shadowers. The blue line indicates the best fit regression line, and the grey band represents the 95% confidence interval.

as many as 4 formant frequencies, but the vowel information is carried primarily by the first two (F1 & F2; Denes, 1995; Hillenbrand, Getty, Clark, & Wheeler, 1995). Following studies that evaluate vowel perception, the current experiment examined F1 and F2 (Hillenbrand & Nearey, 1999). *Vowel dispersion* was used as a measure of variability in baseline vowel production. Vowel dispersion is typically characterized as a measure of speech precision, but can arguably be thought of as a measure of variability. High vowel dispersion is an indicator that the speaker is using a greater portion of the available acoustic-phonetic space and therefore may be less restricted in the ability to produce a range of tokens. Vowel dispersion generally captures the extent to which vowels differ from one another and hence, indexes a type of vocal flexibility that may be associated with vocal alignment.

Vowel dispersion can be computed two ways: on a by-speaker or by-vowel basis. By-speaker vowel dispersion is calculated by taking the Euclidean distance between every individual F1 and F2 value and the central F1 and F2 value for a participant (Bradow et al., 1996). The central value is calculated on an individual basis as the mean of each shadower's F1 and F2, and all values were Labov-normalized prior to analysis. The calculation for by-vowel vowel dispersion is similar, but the center point is the mean F1 and F2 of each *vowel* for a given speaker, rather than a grand mean F1 and F2 for that speaker (Figure 9).

The two measures of dispersion index different aspects of productive variation. A higher numerical value on by-speaker vowel dispersion represents a more dispersed vowel space (e.g., larger percentage of vowel space used) while a smaller value represents a more contracted vowel space. That is, highly dispersed vowels are highly spread out in articulatory space. While high by-speaker dispersion might be an indicator of greater productive variability or flexibility across vowels, it does not capture the amount of variation in the production of specific vowels. By-vowel dispersion captures

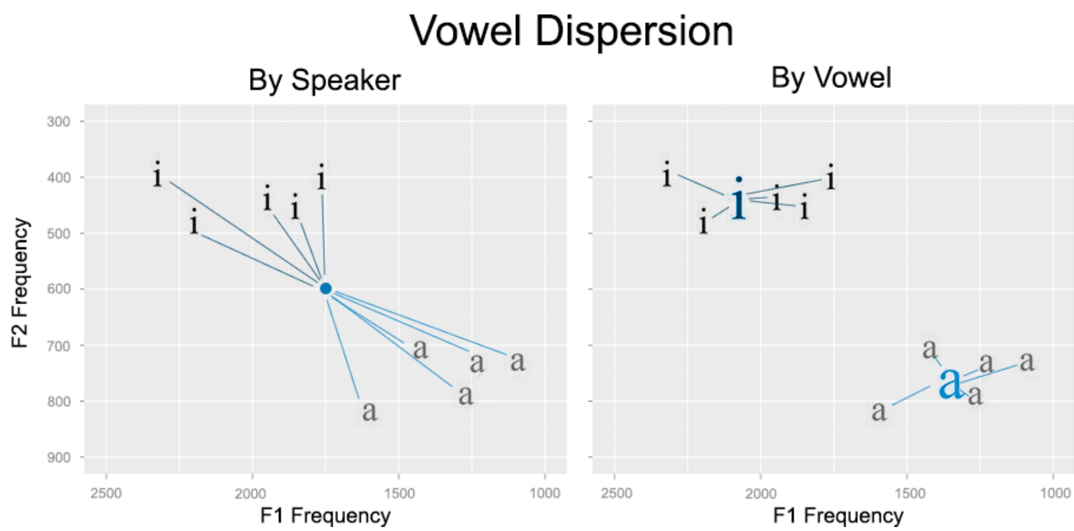

Figure 9. An illustration of the two different ways to compute vowel dispersion: by speaker (left) and by vowel (right). In both cases, a centroid is computed and then values are compared against the centroid. However, whether the centroid is the center of the speaker's vowel space or the center of a particular vowel depends on the measure of dispersion.

variation around any particular token (e.g., variation in the "m*a*ll" vowel) and may index how consistent speakers are when producing a certain vowel class. These two types of dispersion are positively correlated (Figure 10; *R*=.48, *t*(36)=3.31, *p*<0.001), suggesting that they measure related constructs. Speakers with high variability within each vowel category tend to use a larger proportion of available vowel space. If greater variability is associated with more vocal alignment, individuals who are highly variable on both would be expected to exhibit the most alignment to vowel spectra.

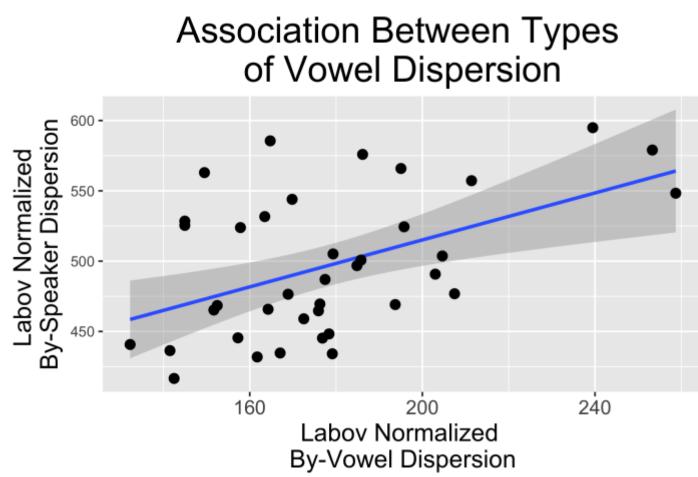

Figure 10. The correlation between by-vowel (x-axis) and by-speaker (y-axis) vowel dispersion is generally positive.

The relationship between vowel dispersion and vowel DID is depicted in

Figure 11. There is a small, positive relationship between both types of dispersion and alignment to vowel spectra. When entered separately into an MEM, neither by-speaker vowel dispersion, $\beta_{spvdisp}$=1.22, *t*=1.22, *p*=.224, nor by-vowel dispersion predicted alignment to vowel spectra, β_{vvdisp}=0.25, *t*=1.61, *p*=.11. Neither type of vowel dispersion reliably improved MEM fit, both $\chi^2(1)<1.8$, *p*>.11. These findings do not support the proposal that greater baseline variability in vowel spectra is associated with larger vowel DID.

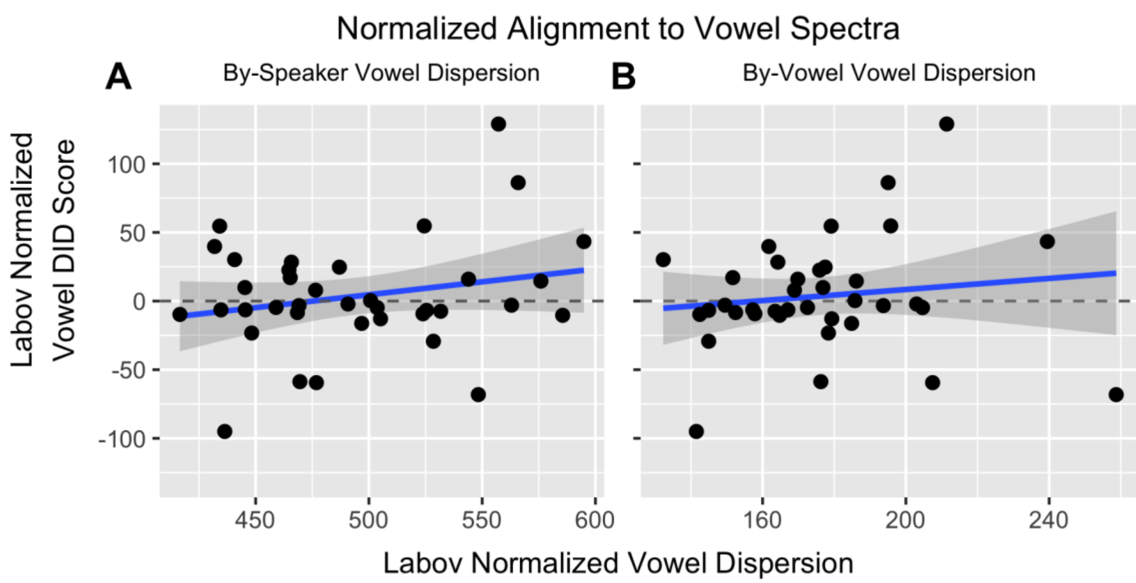

Figure 11. Relationship between by-speaker vowel dispersion (left), by-vowel vowel dispersion (right) and vowel DID. Both panels show a positive relationship between alignment and vowel dispersion. Points represent individual shadowers. The reader should note the difference in x-axis scales across the two types of dispersion

Discussion

The current experiment was designed to test the extent to which vocal alignment required cognitive or attentional resources. First, the degree to which alignment relied on finite cognitive resources (e.g., is efficient) was examined by using a dual-task paradigm where vocal alignment was assessed with and without a cognitive load. The role of attentional or cognitive resources on propensity to vocally align was hypothesized to shed light on the type of mechanism underlying alignment behavior. Second, by using both native- and Spanish-accented speech, the degree to which cognitive load differentially impacted alignment behavior due to differences in social stereotypes and intelligibility was also examined. Differences in shadowers' alignment to speakers of particular language backgrounds can point toward either goal-directed or stimulus-driven properties of alignment behavior. Although others have proposed that aspects of vocal alignment may be automatic (Dragojevic & Giles, 2016; Lewandowski & Nygaard, under review, Pickering & Garrod, 2004), the current experiment marks one of the first direct attempts to examine the automaticity of vocal entrainment processes (Abel & Babel, 2017; Garnier et al., 2013; Solanki et al., 2016).

Automaticity of Vocal Alignment

Efficiency, or the extent to which a process can be initiated without engaging attentional resources (Bargh, 1994; Moors & De Houwer, 2006), is one of the criteria of automaticity. In the current experiment, vocal alignment was indexed by both perceptual and acoustic (duration, fundamental frequency, and vowel spectra) measures. Patterns of alignment were complex, and the effect of cognitive load, subtle. However, when alignment was assessed by human listeners in an AXB task, cognitive load appeared to make no difference to patterns of alignment. This finding suggests vocal alignment may indeed be an efficient process, contrary to the findings of Abel and Babel (2017), who observed a decrease in vocal alignment under increased processing demands.

Furthermore, as indexed by the perceptual task, shadowers aligned marginally more to Spanish-accented model talkers than to native-accented ones, with idiosyncratic patterns of alignment across individual models. If vocal alignment were driven by social preferences, shadowers should have aligned more towards the native targets, who were rated as higher status than the Spanish-accented model talkers. However, the opposite pattern of alignment was observed. The pattern of results suggests that, in a context that minimizes explicit social constraints, vocal alignment may function to increase intelligibility of speech and highlights the role of low-level perceptual-motor couplings (Adank et al., 2010; Lewandowski & Nygaard, under review; Shockley, Sabadini, & Fowler, 2004).

Cognitive Control of Vocal Alignment

Although listeners did not perceive a difference in alignment across cognitive load conditions, an effect of load emerged in the acoustic measures. That an effect was present but not perceptible to raters suggests any influence of load on alignment was subtle. For both duration and F0, shadowers exhibited greater vocal alignment in the CL condition, but there was no effect of model talker's accent. Because vocal alignment to these acoustic parameters increased when participants engaged in a cognitively demanding secondary task, this may be evidence that attention or cognitive resources work to modulate an underlying, largely automatic process. When those modulatory resources are impacted, vocal alignment increases (i.e., van Baaren et al., 2009). This is consistent with findings of imitation in the perceptual-motor domain (van Leeuwen et al., 2009), which show facilitation of behavioral alignment when cognitive demands are high.

By contrast, shadowers exhibited attenuated alignment to vowel spectra under cognitive load, leading to the apparently paradoxical finding that cognitive load both increases and decreases alignment on acoustic parameters. A key difference between F0 and duration and vowel spectra may account for this conflict: F0 and duration are not explicitly linguistic acoustic dimensions, but vowel spectra are. Fundamental frequency (pitch) and speaking rate (duration) are supra-segmental properties that are present across a variety of vocalization types. These features are also arguably more variable within an individual, and can be altered quite rapidly. By contrast, vowel spectra point specifically to vowel quality and provide linguistically relevant acoustic information. Vowel spectra may be more stable across time because of their direct contribution to conveying propositional content. The distinctions between suprasegmental and segmental properties of the speech signal may have consequences for how they are encoded and how they can change under high processing load and as a function of accent.

Recent work suggests that cognitive load may impair an individual's ability to process fine-grained linguistic structure, particularly in phoneme identification (Mattys & Wiget, 2011). For example, Mattys, Barden, and Samuel (2014) tasked participants simultaneously completing a phoneme discrimination task and a visual search task of varying difficulty. As cognitive load increased, participants showed an increased reliance on lexical context to make phoneme discrimination judgments, suggesting that they were attending less to the actual fine-grained acoustic-phonetic structure in the speech signal. In the current experiment, the overall lack of alignment to vowel spectra in the cognitive load condition may be due to a decreased perceptual acuity caused by the load (Mattys & Palmer, 2015). However, fine-grained linguistic discrimination ability may not be necessary for alignment to F0 and duration. Mattys and colleagues' work explores discrimination of explicitly linguistic characteristics of speech, such as words or phonemes (e.g., speech sounds). Because F0 and duration are supra-segmental characteristics of speech and may represent more general properties of voice and emotion, they may not be subject to the same level of perceptual impairment under load.

Relationship Among Measures of Vocal Alignment

Across the acoustic measures used to assess vocal alignment in the current experiment, patterns of alignment were not uniform. Particularly with respect to the effect of model talker's accent, the perceptual measures suggested that shadowers aligned more to Spanish-accented talkers while the vowel spectra suggested that shadowers in the no load condition showed less alignment (or even divergence) on vowel spectra DID. Although the acoustic dimensions examined in the current experiment have been shown to influence perception of vocal alignment (Lewandowski & Nygaard, under review; Pardo et al., 2013b), they represent only subset of the the many dimensions along which alignment and divergence may occur (Bilous & Krauss, 1988; Pardo et al., 2013b). Raters in the AXB task may be attending to or sensitive to changes that were not captured by the acoustic measures reported here, which could account for the differences in vocal alignment observed across measures.

Individual Differences in Speech Production

In addition to providing evidence of automaticity in vocal alignment, the current experiment offered the opportunity to replicate and extend previous findings in the literature regarding individual differences that may contribute to alignment. Most researchers define a language user's production repertoire in terms of dialect features

(Babel, 2012; Walker & Campbell-Kibler, 2015; Kim et al., 2011; but see Lewandowski & Nygaard, under review). Here, shadower's speech production tendencies were defined in terms of variability or flexibility at the individual level. The current experiment found positive but non-significant associations between baseline production variation in F0 and vowel spectra and acoustic alignment, similar to the findings of Lewandowski and Nygaard (under review). In addition to dispositional differences, variation in speech production abilities (Lewandowski, 2009) may be an individual difference related to propensity to align.

Summary

The results of the current experiment suggest that (1) vocal alignment does not require cognitive or attentional resources, (2) attentional resources may serve to modulate vocal alignment under normal processing demands, (3) the role of baseline variation in determining individual vocal alignment behavior is weak. Although listener judgments of vocal alignment suggested that cognitive load had little impact on degree of alignment, load did influence alignment to acoustic properties of speech. However, overall, the results of the current experiment imply that vocal alignment behavior is efficient. With respect to the role of cognitive resources, alignment on acoustic measures suggested that attentional resources may be dedicated to modulating degree of vocal alignment. The findings of Experiment 1 highlight the roles of both low-level cognitive and perceptual mechanisms and high-level cognitive control in driving vocal alignment behavior. These findings support reframing our understanding of vocal alignment in terms of the automaticity of underlying processes rather than exclusively in terms of perceptual-motor links (e.g., Interactive Alignment Model, Pickering & Garrod, 2004; 2006) or of general socio-motivational goals (e.g., Communication Accommodation Theory, Giles et al., 1991; Soliz & Giles, 2014).

However, patterns of alignment differed across model talkers of native and non-native language backgrounds in a manner that largely followed the acoustic-phonetic differences rather than the social information. An exception to this general pattern was found in alignment to vowel spectra, where English models received more alignment than Spanish models from shadowers in the no load condition. This difference disappeared under cognitive load, perhaps because of the nature of alignment to fine linguistic structure (Mattys et al., 2014, Mattys & Wiget, 2011) or the degree of familiarity with the realization of Spanish versus English vowel categories (Bradlow, 1995; Munro & Derwing, 1995). Overall, social information associated with the model talkers were not as predictive of patterns of vocal alignment as differences in intelligibility. This could be due in part to the type of social information involved. In other words, Experiment 1 relied on stereotypes generally associated with groups of speakers, but these may have been overshadowed by differences in intelligibility. Experiment 2 assesses whether another kind of social information, an inferred personality trait, may modulate vocal alignment behavior and the extent to which this may be resource-demanding.

Experiment 2

Previous research indicates that socially relevant information, such as facial and verbal cues to emotion, is rapidly encoded and used to guide behavior (Hansen & Hansen, 1988; Schulz, Mothes-Lasch, & Straube, 2013; Vuilleumier et al., 2004). Infants are able to detect helpful and harmful social behavior as early as 6 months of age (Hamlin, Wynn, & Bloom, 2007). Adults readily integrate both verbal and non-verbal signals of trustworthiness, and can use that information, for example, to predict whether an individual will be a trustworthy partner in economic games (DeSteno et al., 2012). These examples point to an individual's direct experience and assessment of others. However, an individual's assessment of others may also come from indirect sources. Stereotypes associated with an individual's group membership and secondhand accounts (i.e., gossip) about an individual's actions can likewise have consequences for behavior. The current experiment explores the role that character attributions may play in vocal alignment.

Social Stereotypes

Stereotyping is a form of categorization that allows individuals to call upon associations between groups and common traits. This process allows for simple, heuristic assessment of others, even in the absence of empirical evidence to make judgments (Allport, 1954; Devine, 1989; Tversky & Kahneman, 1973). The activation of person-associated stereotypes and social constructs more generally have consequences for behavior. For example, the presence of an Asian-American experimenter elicited more stereotype-consistent word completions (e.g., S_Y being completed as SHY rather than STY or SPY; Gilbert & Hixon, 1991). Men primed to believe that a female they were about to meet was attractive were more likely to rate her as attractive and sociable (Snyder, Tanke, & Berscheid, 1997). Even traits that are not specifically attributed to a person but to social constructs associated with group membership more generally can influence behavior.

For example, participants in the first experiment of Bargh, Chen, & Burrows (1996) completed a scrambled sentence test (e.g., Srull & Wyer, 1979) in which subjects must use five scrambled words to compose a grammatical four-word sentence. The task's purpose was ostensibly to test language ability. Critically, adjectives and adverbs in the scrambled sentences included concepts related to rudeness, concepts related to politeness, or neutral concepts. After completing the task, participants encountered a situation in which they could interrupt a conversation (e.g., behave rudely) or wait for the conversation to finish (e.g., behave politely). Participants who had been primed with polite constructs were significantly more likely to wait the full 10 minutes for the conversation to end than participants in either the rude or neutral priming conditions (but see Doyen, Klien, Pichon & Cleeremans, 2012).

A number of researchers have demonstrated links between social representations and actions and adjustment of behavior, including behavioral alignment to others (Chartrand & Bargh, 1999; Dijksterhuis & van Knippenberg, 1998). Social information may have similar consequences for alignment to speech characteristics. Consistent with this proposal, attitudes towards an individual's group (i.e., nationality, sexual orientation) influences the degree of vocal alignment to a perceived member of that group (Abrego-

Collier et al., 2011; Babel, 2010; Yu et al., 2013). Using a laboratory naming task, Sidaras (2011) examined vocal alignment to an age-ambiguous male talker who spoke in a neutral speaking rate. Prior to shadowing the model talker, participants read a narrative that described “Mr. Jones,” a 70-year-old retiree who now lived in Florida, or “Tommy,” an adaptable 22-year-old who moved from New York City. Although the same, neutral-rate (i.e., neither fast nor slow) utterances were target stimuli in both conditions, shadowers who had seen the Mr. Jones description slowed their speaking rate in the subsequent shadowing block while shadowers who had seen the Tommy description accelerated their speaking rate. Shadowers engaged in “illusory” alignment; that is, shadowers aligned to speech characteristics that they expected due to the model talker’s supposed social category.

Secondhand Experience of Others

Even in the absence of direct information about social attributes associated with groups, individuals infer characteristics, draw conclusions, and make judgments of others. These inferences in turn affect behavior and memory (Carlston & Skowronski, 1994; 2005; Todorov & Uleman, 2002). For example, Winter and Uleman (1984) examined memory for actions and agents based on declarative sentences that contained diagnostic cues to an agent’s disposition (i.e., “The professor has her new neighbors over for dinner,” “The reporter steps on his girlfriend’s feet as they foxtrot”). The authors found that dispositional cues implied by such statements (i.e., “friendly,” “clumsy”) were more effective in promoting recall for sentences than semantic associates of other objects in the sentence, even though the dispositional cues themselves were not present at the time of encoding. Despite the success of trait labels in cuing recall, participants were unaware of having made trait inferences at all. This finding suggests that social evaluation occurs rapidly and that social judgments have consequences for memory and behavior.

Information about the actions of individuals recounted through short narratives may also penetrate low-level perceptual processing of socially relevant stimuli. For example, affectively neutral faces paired with sentences describing positive or negative behavior were subsequently rated as being more positive or negative relative to faces paired with neutral sentences (Bliss-Moreau, Barrett, & Wright, 2009). Anderson and colleagues (2011) paired affectively neutral faces with positive, negative, or neutral statements that were either social (i.e., “helped an elderly woman with her groceries,” “threw a chair at a classmate”, “passed a man on the street”) or non-social (i.e., “felt the warm sunshine,” “had a root canal,” “drew the curtains in the room”) in content. The faces were then presented against each other or against images of houses in a binocular rivalry task (one image presented independently to each eye). The researchers found that faces that had been previously paired with negative social statements were consciously experienced and reported more of the time in the binocular rivalry task than images of houses or of faces that had been paired with positive or neutral social statements.

The authors argue that the faces paired with negative statements became more threatening, and thus dominated perception. These findings indicate that social inferences can penetrate low-level perceptual processes and demonstrate one mechanism by which attention to social information may direct behavior. If the same social representations that affect perception and memory affect cognition more generally, then the same effects may also be present in vocal alignment. Although not the authors' main goal, Bourhis and Giles' (1977) study demonstrated that direct observation of an interviewer's disposition (e.g., rude, unfriendly) altered alignment behavior. A speaker of Received Pronunciation (RP; a high-status dialect of British English) interviewed Welsh-speaking adults. Interviewees tended to converge toward the interviewer until they "overheard" (this was a staged event) the interviewer disparage Welsh language and culture. Participants who identified strongly with their Welsh heritage subsequently adopted more pronounced Welsh accents. One participant even conjugated Welsh verbs instead of responding to the remaining interview questions. Considering the overt insult and subsequent behavior of the interviewees in this task, it suggests that social information can result in deliberate tuning of alignment behavior.

Given the nature of an interview setting, any number of variables may have been at play. One aim of the current experiment was to assess the importance of trait inferences in modulating entrainment to speech in a setting that minimizes the influence of communicative and situational constraints. A further goal was to determine whether such information could serve as an inhibitor of vocal alignment when it points to socially deviant or undesirable traits. The manner in which social information interacts with vocal alignment may be in part conditioned on the automaticity of trait inference itself.

Automaticity of Social Evaluation

Recalling Bargh's (1994) criteria for automaticity, trait inference from short narratives like the ones used by Uleman and colleagues (Todorov & Uleman, 2002; Winter & Uleman, 1984) seems to occur without the individual's awareness or intent

(Lupfer et al., 1990; Winter et al., 1985). However, trait inferences are subject to contextual and processing demands; for example, trait inference may not be as spontaneous for members of certain cultures or for individuals who are less prone to emphasizing the self (Na & Kitayama, 2011; Duff & Newman, 1997). Furthermore, working memory load has been shown to decrease the degree to which trait inferences are made (Wells, Skowronski, Crawford, Scherer, & Carlson, 2011; Uleman, Newman & Winter, 1992), suggesting that trait inference requires processing resources. This is in contrast to social stereotype priming, which suggest that activation of social constructs may occur without awareness or intent, particularly for stereotypes that broadly permeate the cultural consciousness (Bargh, 1982; Chen & Bargh, 1997; Devine, 1989). Trait inferences are therefore another potential mechanism by which cognitive processes may modulate vocal alignment.

Social Evaluation and Vocal Alignment

Although past research clearly demonstrates a role for social information in vocal alignment, there are lacunae in the understanding of how such information relates to vocal alignment that must be addressed. First, the types of social information examined, in the form of stereotypes, attitudes, and trait inferences may differ from one another. In previous work examining the influence of social attitudes on vocal alignment, stereotypes are frequently used as a proxy for or as an inducer of attitudes (Experiment 1; see also Abrego-Collier et al., 2011; Babel, 2010; Sidaras, 2011). Stereotypes are generated by association, rather than by inference (Bargh, 1984, 1994; Wells et al., 2011; Wigboldus, Dijksterhuis, & van Knippenberg, 2003). Social stereotypes have also been found to interfere with an individual's ability to make trait inferences if the implied trait is inconsistent with the stereotype, particularly when processing demands are high (Ramos, Garcia-Marques, Hamilton, Ferreira, & van Acker, 2012; Wigboldus, Sherman, Franzese, & van Knippenberg, 2004). These findings suggest that prioritization and use of information gleaned through different channels may qualitatively differ.

Second, whether social information requires processing resources to guide behavior (i.e., is efficient) has important implications with respect to the interface between social information and vocal alignment. The answer to this question may depend on the type of evaluation being queried, i.e., if a stereotype, a secondhand account, or direct experience is the source of social information. Social stereotype activation is hypothesized to rely on automatic perception-production links (Bargh, 1982; Chen & Bargh, 1997; Devine, 1989) consistent with general principles of perception and action (see Chapter I, Cognitive and Perceptual Mechanisms Supporting Vocal Alignment), while trait inference seems to involve an intermediate step of translating social information to behavior (Chartrand, Maddux, & Lakin, 2005). In other words, inferring a trait from and altering behavior in response to secondhand accounts of social information may involve either implicit or explicit decision-making. By contrast, activating a social stereotype occurs effortlessly, even if using social stereotypes to guide behavior may require more attentional or cognitive control (Culter, 2010; Devine, 1989). Thus, sources of social evaluation may rely on different underlying mechanisms and have divergent consequences for alignment.

Overview of Experiment 2

The goal of Experiment 2 was to assess the mechanism by which inferred social traits influence vocal behavior. Specifically, Experiment 2 sought to evaluate (1) the extent to which inferred positive or negative trait attributions influence vocal alignment and (2) whether this process is efficient. In Experiment 2, short narratives from which a dispositional inference about each model talker could be made preceded shadowing trials in a dual-task paradigm like that used in Experiment 1. In order to limit the negative stereotypes typically associated with speakers of non-standard English and capitalize on the effect of trait ascriptions, only native English speakers served as model talkers. It is expected that the valence of inferred traits will affect vocal alignment behavior. That is, shadowers will align more to talkers who are more likeable (i.e., paired with positive information) and align less to talkers who are less likeable or unlikeable (i.e., paired with negative information).

As in Experiment 1, adding cognitive load may reveal the extent to which using trait inference to guide behavior does or does not require processing resources (i.e., is efficient). Depending on the efficiency with which social information influences vocal alignment, the dual-task naming paradigm should reveal different results. If trait attributions can be made and used to direct behavior in an efficient, non-resource demanding way, the effect of social information should be identical across load conditions. This would be consistent with a social priming type mechanism (Chartrand & Bargh, 1999; van Baaren et al., 2009). If using socially evaluative information to direct alignment requires attentional resources, then individuals under cognitive load should align to both talkers equally, regardless of the talkers' inferred valence.

Prior to executing Experiment 2, a preliminary experiment was conducted to establish that social information as provided by photos and narratives would influence vocal alignment as indexed by speaking rate (duration). In this preliminary experiment, utterances were duration manipulated to be “fast” or “slow” using Praat (Boersma & Weenink, 2014) with a procedure that altered perceived speaking rate without affecting vocal pitch. Modifying duration in this way introduced variation into speaking rate, the dimension of interest. Forty-seven individuals recruited from the Emory University Psychology subject pool (ages 18-25) participated in a shadowing task similar to the one used in Experiment 1. Prior to shadowing, however, participants saw a black and white photo of a face and a short story describing the experimenter’s “good” or “bad” experience with that person. Shadowers generally approximated the specific speaking rate used by the model talker (e.g., duration difference in distance) and altered their speaking rate in the same direction as the model talker (e.g. duration change direction). Importantly, alignment to duration differed by social information, suggesting that the introduced social information may influence vocal alignment behaviors. Building on the evidence for speaking rate serving as an index of general vocal alignment found in both Experiment 1 and the preliminary experiment, Experiment 2 used simplified the analysis procedures by limiting assessment to alignment on duration and examined the effect of cognitive load and social information on vocal alignment to duration.

Methods

Participants

All participants were recruited from the Emory University psychology subject pool. As in Experiment 1, participants were 18-25 years of age, native speakers of American English, and had no self-reported history of hearing or speech disorders. Specific exceptions are noted below. Furthermore, participants in Experiment 2 had not participated in Experiment 1. All shadowers received partial credit towards a psychology course requirement.

Shadowers.

Sixty individuals (23 male, 37 female) were recruited to participate in the naming task. Thirty were randomly assigned to the cognitive load (CL) condition and 30 to the no load (ND) condition. Sixteen shadowers were excluded from data analysis due to computer technical error (6), failure to follow task instructions (5), failure to complete the experimental task (2), history of hearing or speech disorder (2), or failure to complete the EQ (1). Of the remaining 44 shadowers, equal numbers of participants had been assigned to the CL (13 female, 9 male) and ND (17 female, 5 male) conditions.

Materials

Word stimuli.

A list of 96 bisyllabic words was constructed from items in the English Lexicon Project database (Balota et al., 2007). This word list was made to be balanced on syllable stress (first vs. last), word frequency (high vs. low), stressed syllable onset (voiced vs. voiceless consonant; fricative vs. nasal vs. plosive manner of articulation), and stressed nucleus vowel (/i/, /ɛ/, /ɑ/, /u/). The median frequency for items classified as "low frequency" was 215.5 occurrences per 131 million ($\bar{x}$=620.23; Burgess & Livesay, 1998), and the median for high frequency words was 37,380.5 occurrences ($\bar{x}$=72,945.33)[4].

Model talkers.

Model talkers were chosen on the basis of the talkers' familiarity with the word list, attribute ratings, and perceptual discriminability of their voices. Fifty native English speakers were recorded as they read the 96-item word list. Of these 50, three native English-speaking males were selected as candidate model talkers. Utterances recorded by the potential model talkers were presented to 19 independent listeners and assessed on 21 attribute pairs (e.g., nice-mean, honest-dishonest, arrogant-humble, smart-dumb; Heaton & Nygaard, 2011) on a 7-point Likert-type scale. Whether positive or negative attributes were anchored to "1" or "7" was counterbalanced across participants. The 21 attributes were reverse scored so that positive numbers always indicated more of the positive attribute and were then grouped into two categories ("sociality" and "solidarity") using a k-means clustering approach. The two most similarly rated talkers were

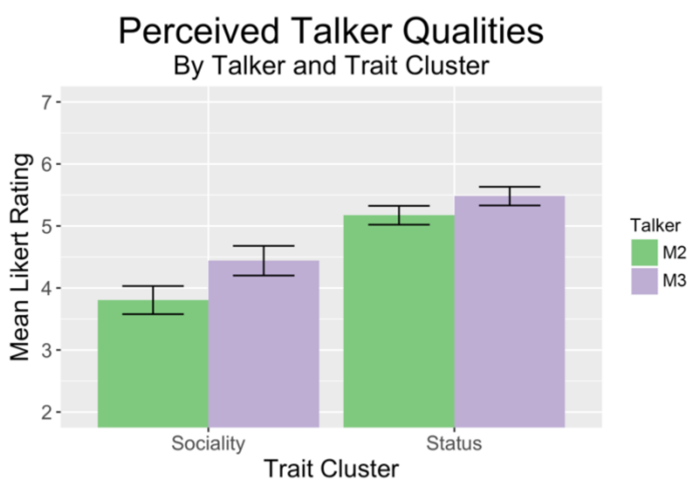

Figure 12. Ratings on the two model talkers, M2 (green bars) and M3 (lavender bars), grouped by trait cluster. Error bars represent standard error of the mean.

[4] Word frequency was included in analyses as a fixed effect factor of interest. It was found that alignment to duration was significantly greater on low frequency items than on high frequency items (β_{low}=7.99, t=2.38, p=.0193). However, word frequency did not interact with the primary effects of interest (cognitive load, social information; all β<3.5, t<.587, p>.5575) and will not be discussed further.

initially selected as target speakers. However, these two voices were difficult for pilot participants (n=3) to discriminate, so M2 and M3 were selected as final model talkers. Mean ratings for these two talkers are shown in Figure 12.

Utterances were duration manipulated in Praat (Boersma & Weenink, 2014) to introduce variation into the acoustic dimension of interest. Fast ($\bar{x}_{M2}$=409.72 ms, $\bar{x}_{M3}$=430.92 ms) and slow ($\bar{x}_{M2}$=668.02 ms, $\bar{x}_{M3}$=798.51 ms) versions of each item relative to the shadower's own neutral speaking rate were created. The algorithm manipulated duration while holding fundamental frequency (pitch) constant. The manipulated utterances sounded natural, though some shadowers did report that the model talkers 'dragged their words out' for the long duration stimuli.

N-back image stimuli.

The Fribble images used in Experiment 1's visual 1-back task were used again here.

Character narratives.

To manipulate social information, two narratives were chosen from a subset of unpublished character narratives (Beal, 2018). Each narrative was written in simple, declarative sentences and had between 31 and 44 words ($\bar{x}$=35). Independent raters judged the extent to which each narrative described the attractiveness, character, cheating, confidence, competence, trustworthiness, and warmth of its protagonist as well as the intensity, valence, and weirdness of the narrative itself on a 7-point Likert-type scale (Appendix B). The following criteria guided selection of the two narratives: (1)

rated low on weirdness, (2) described someone who could plausibly be an Emory University student, and (3) were similarly rated across the non-critical attributes. One narrative described a “generous/warm” individual while the other described a “vicious” individual. These will hereafter be referred to as the “good prime” and “bad prime,” respectively. This nomenclature is adopted for narrative expediency only and does not reflect endorsement of a social priming account of vocal alignment.

Character photos.

Two black-and-white photos of Caucasian male faces with affectively neutral expressions were chosen to accompany the written character narratives (Figure 13). The photos were selected from the Productive Aging Laboratory (PAL; Minear & Park, 2004) Face Database on the basis of ratings collected by Ramos et al. (2016). The Ramos et al. rating battery included the dimensions of attractiveness, dominance, gender typicality, likeability, perceived age, and trustworthiness. When averaged across Ramos et al.’s 223 participants, the two chosen photos differed from each other by less than half a point on any dimension (max. difference=.43, ratings on a 7-point Likert-type scale), making

Figure 13. The face images used alongside the character narratives to provide social information.

them more similar than any photo pair in the set. Ramos et al.'s (2016) ratings on these face stimuli are provided in Appendix B.

Empathy Quotient.

As in Experiment 1, the Baron-Cohen and Wheelwright (2004) EQ was administered to shadowers. In order to achieve a sufficient sample size for an analysis of individual differences in disposition, the empathy data from Experiment 1, 2, and the preliminary experiment were pooled for analysis.

Procedure

Naming task.

The naming task used in Experiment 2 is nearly identical to the procedure used in Experiment 1, with some differences. As in Experiment 1, shadowers were instructed to verbally name (i.e., say aloud) visually printed words (baseline block) or aurally presented words (shadowing block). Given the range of performance on the 1-back in Experiment 1, it was unclear whether or not the shadowers had sufficient practice with the 1-back before beginning the experimental trials. To take into account potential individual differences in task learning rate, participants were given a minimum of 20 and a maximum of 60 practice trials to reach 80% accuracy on the 1-back task, regardless of whether they were in the CL or ND condition. The baseline block followed practice and was exactly as described in Experiment 1 (see Figure 14).

Unlike Experiment 1, a perceptual exposure block was not included in Experiment 2. According to the social priming account described in the introduction

(Bargh, 1982; Chartrand et al., 2005; Chen & Bargh, 1997), hearing a talker's voice should automatically prime social stereotypes. In order to minimize the likelihood that participants would implicitly make social judgments about the model talkers before encountering the shadowing block, the perceptual exposure block was not included. Furthermore, both model talkers were native English speakers, eliminating the need for participants to attain familiarity with the accent.

Prior to beginning the shadowing block, participants were presented with both character primes (photo and written narrative) presented sequentially. The primes were couched in terms of the experimenter's experience: the narratives were presented in first person singular and indicated that the experimenter herself had seen the behaviors described in the narratives from the model talkers. Participants had unlimited time to read about 'my experience' with each person before moving to the next, but could not revisit a prime once they had finished reading it. Participants advanced to the next narrative or to the next phase of the experiment by pressing the space bar. Primes were presented in a random order, and the photo-prime-model talker pairings were counterbalanced across participants using a Latin-square design.

Note that at the time the shadowers read the character primes, they had not yet heard the model talkers' voices. Because the utterances were not blocked by speaker, participants needed to be able to map the appropriate voice to each name during the shadowing block. To provide a cue to talker identity, the photo initially presented with the prime co-occurred on each trial with the corresponding model talker's utterance. The photo remained onscreen until shadowers initiated their naming response or for a maximum of 2000 ms. In all other respects, the timing and presentation of speech and 1-

back stimuli were identical to Experiment 1. Shadowers heard an equal number of fast and slow utterances spoken by each model talker, and each talker produced a unique subset of the 96 words (48 per model talker, counterbalanced across participants).

Because Experiment 2 posed greater perceptual demands (2 similar English-speaking male targets vs. 1 target per distinct gender/accent groups) and memory demands (good vs. bad prime) on participants, the participants were given a surprise talker identification task after the conclusion of the shadowing block. The identification task consisted of 10 trials per model talker. Half the words presented were ones that had been spoken by the opposite talker in the shadowing block and half were produced by the same talker as during shadowing. On each trial, a word was played to the shadower, and then participants pressed a key to indicate their response. The photos and names associated with each voice were shown on the decision screen. Participants did not receive feedback.

Upon completion of the talker identification task, shadowers were prompted to answer five questions: (1) if they had a preference for one of the talkers (yes or no), (2) if yes, which talker did they prefer and why (in 200 characters or less), (3) if yes, how strong was their preference, (4) which person did the experimenter have a positive experience with, and (5) which person did the experimenter have a negative experience with. After the answers to these questions were recorded, shadowers filled out the Empathy Quotient. Except for the EQ, which was administered on paper, all testing procedures were administered on a Windows PC running Windows XP in Eprime 2.0 experiment software (Psychology Software Tools, Inc., 2002).

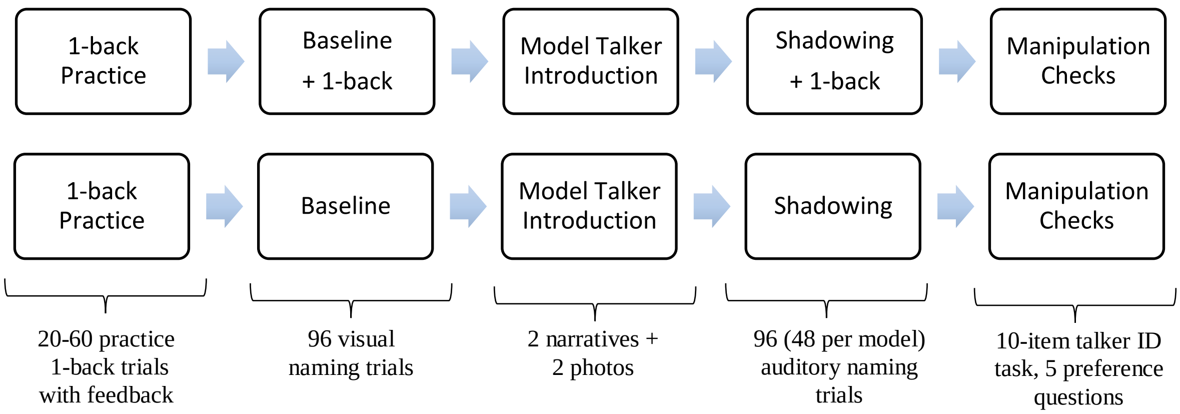

Figure 14. Schematic figure of Experiment 2's task structure. Top panel depicts the structure for CL conditions, and the bottom illustrates the paradigm for ND conditions. Model talker pairings with the social information (narrative and photo) and presentation order were counterbalanced across participants.

Results

As in Experiment 1, the data were analyzed using MEM regression techniques. Duration difference in distance (DID) was the primary measure of interest, and analysis was carried out with general linear mixed-effect models (GLMM). For all MEMs, the maximal random effects structure justified by the sample was used[5] (Barr et al., 2013). Categorical predictors were deviation coded (-.5, .5) and the significance of fixed effects was justified by within-MEM tests of significance and cross-MEM comparisons. Prior to analysis, trials associated with shadowed naming RTs > ±2 SDs away from a participant's mean were excluded. This trimming procedure reduced the available data by 4%. Mispronunciations of the target accounted for another 7.86% of trials. Due to participants' low familiarity with the low frequency words in the word set, mispronunciations were not excluded unless the intended shadowed word was completely

[5] The maximal RE structure that converged for all MEMs was (1+prime | shadower) + (1 | word).

unclear or a different number of syllables than the target (e.g. “resumé” for “resume”). Multi-collinearity among nominal predictors (0.006 < Cramér’s V < 0.15), nominal and numeric predictors (-0.0005 < ICC < 0.18), and numeric predictors (11.87 < κ < 28) was low to moderate. Where relevant, specific exceptions are noted in the sections below.

Preliminary Analyses

N-back task familiarization.

The majority (42/44) of shadowers achieved a mean accuracy of 92.85% in 20 trials during the 1-back practice administered prior to the experimental tasks (see Figure 14). The two shadowers who needed more than 20 trials took 60 (CL condition) and 25 (ND condition) trials to achieve 80% accuracy. Because the majority of shadowers were able to achieve criterion performance within 20 trials, this suggests that the 20-trial practice of Experiment 1 was sufficient.

Voice discrimination and talker preferences.

In order to form an explicit talker preference based on the social prime, shadowers must (1) be able to accurately tell the two talkers apart, (2) read the primes when they are presented, and (3) remember which prime was associated with which voice. These preconditions as well as the success of the prime at driving preferences were assessed.

Shadowers discriminated between the talker’s voices with 96.5% accuracy. There were no statistically significant differences between the CL condition and the ND condition on either talker identification accuracy (t(42)=1.44, p=.158) or talker identification response time ($t(42)=-.33$, $p=.743$). Participants in both conditions were equally able to perceptually discriminate and correctly remember the talker's voices.

On average, shadowers spent approximately half a minute ($\bar{x}$=26.64s) reading the primes. Because the presentation order was fully randomized, 27 subjects saw the good prime first and 17 saw the bad prime first. There was a significant prime (good vs. bad) x presentation order ('good first' vs. 'bad first') interaction on reading time, β=6569, $t=3.41$, $p=.001$. The estimate indicates that shadowers spent longer reading the good prime, but only when the good prime was presented first ($\bar{x}_{\text{first}}$=21.24s; $\bar{x}_{\text{second}}$=16.08s). There was no difference in reading time for the bad prime ($\bar{x}_{\text{first}}$=14.89s; $\bar{x}_{\text{second}}$=13.49s). The difference observed cannot be attributed to story complexity, as the bad prime was more difficult to read (Flesch Reading Ease, FRE=49.2; Flesch, 1948) than the good prime (FRE=86.6). The good prime did contain more words than the bad prime (good=44 words; bad=36 words) and more subjects were randomly assigned to the 'good first' presentation order. The main result of interest is that shadowers were attending to the narratives, as indexed by the amount of time they spent on each narrative.

The next question of interest was whether or not shadowers remembered which model talker was associated with each prime. As a whole, shadowers were 78.31% accurate at recalling which talker had been presented as good and which talker had been presented as bad. Thirty-three shadowers correctly identified the prime associated with each talker. Eight shadowers reversed the good/bad mapping of the two talkers. Of the these shadowers, six were in the CL condition and two were in the ND condition. Three shadowers correctly identified one talker as good or bad, but not the other, and two of these shadowers were in the ND condition. Taken together with the voice discrimination results, the majority of shadowers appeared able to form preferences based on the primes for the correct talkers. However, to ensure that vocal alignment was assessed for the correct prime-talker pairings, for the analyses of alignment by prime, the reported association between prime and talker was used.

Just over half of subjects (n=26) reported an explicit preference for one talker over the other. The remaining subjects did not indicate a preference. One shadower indicated a preference for one talker's voice but the other talker's character; one shadower indicated a preference but did not report a talker's name. For shadowers who expressed a clear preference (n=24), the preferred talker was much more likely to be the one associated with the good prime than with the bad prime, $\chi^2(1)=6.75$, $p=.009$. Although not all shadowers explicitly stated a talker preference, the data indicate that shadowers generally formed preferences based on the social character primes.

Because cognitive load condition was equally represented among the shadowers who reported an explicit talker preference, the effect of load condition on expressed preference was also examined. Shadowers in the ND condition formed their preferences largely based on prime, $\chi^2(1)=13.50$, $p=.0002$, but shadowers in the CL condition did not, $\chi^2(1)=0$, $p=1$. The apparent lack of preference formation in the CL condition may have been due in part to the participants who reversed the mapping between the good/bad talkers. Three of the six shadowers in the CL condition who misreported which talker was "good" and which was "bad" also explicitly reported a talker preference. All three of these shadowers expressed a preference for the bad talker, who they misattributed as the good talker. Recategorizing these individuals by their remembered association, shadowers in the CL condition formed their preferences based on which talker they associated with the good prime, $\chi^2(1)=4.17$, $p=.041$. These findings suggest both that keeping the correct mapping in memory was impacted by the CL task and that preferences in general followed mappings to good primes.

Trait Attributions and Cognitive Load

Duration DID score was calculated according to the equation laid out in Experiment 1. By this measure, positive numbers indicate vocal alignment and negative numbers indicate divergence (e.g., shadowers' vocalizations became less similar to the model talkers' utterances). The magnitude of the DID indexes the degree of alignment or divergence. Although multi-collinearity was overall low, shadower was perfectly associated with cognitive load condition (CL or ND; Cramér's V=1) and with model talker preference (ICC=1), both of which were fixed effects of interest in the current set of analyses.

Prime and cognitive load.

Inspection of the means (see Figure 15) suggests that shadowers aligned more to the good talker than the bad in the ND condition ($\bar{x}_{good}$=23.40 ms; $\bar{x}_{bad}$=14.98 ms), but this pattern was reversed in the CL condition ($\bar{x}_{good}$=17.62 ms; $\bar{x}_{bad}$=25.61 ms). There was no

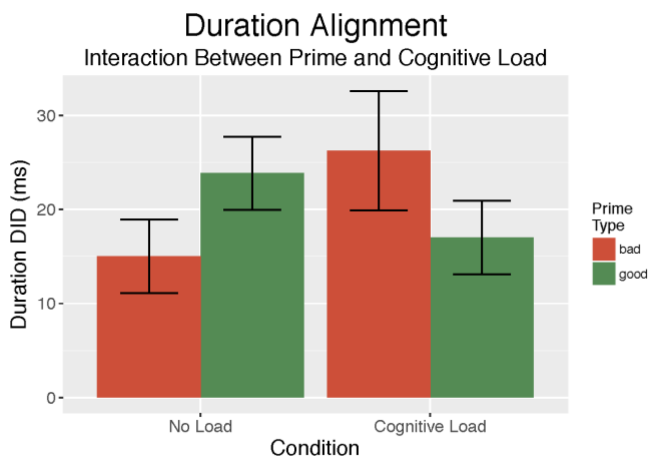

Figure 15. Interaction between cognitive load condition and character prime on vocal alignment to duration. Unless otherwise noted, all error bars represent standard error of the mean.

significant main effect of either prime (β_{good}=-.35, *t*=-.08, *p*=.934) or of cognitive load (β_{CL}=-.91, *t*=-.187, *p*=.853) on duration DID. Neither prime nor load independently improved MEM fit, both χ^2<.031, *p*=.86. However, the load x prime interaction was statistically reliable, $\beta_{CLxgood}$=-18.1, *t*=-2.27, *p*=.028. The interaction term improved MEM fit over a model that excluded it, $\chi^2(1)$=4.86, *p*=.028. To follow up on the interaction, the data were split by condition (CL vs. ND) and the effect of prime was reassessed while condition was held constant. Shadowers in the ND condition exhibited differential alignment based on prime (β_{good}=9.04, *t*=2.29, *p*=.033), whereas shadowers in the CL condition did not (β_{good}=-7.92, *t*=-1.17, *p*=.257). These results indicate that shadowers vocally aligned more to the voice paired with the good prime in the ND condition, but shadowers did not differ in alignment across talker primes in the CL condition. It should be noted that shadowers in both conditions aligned to both talkers; no divergence from the bad talker was observed.

Photo and prime.

Although care was taken to choose photos that were equated on perceived traits, alignment as a function of face traits was also assessed. Shadowers aligned equally to both talker photos, β=-.395, *t*=-.10, *p*=.922. The photo x prime

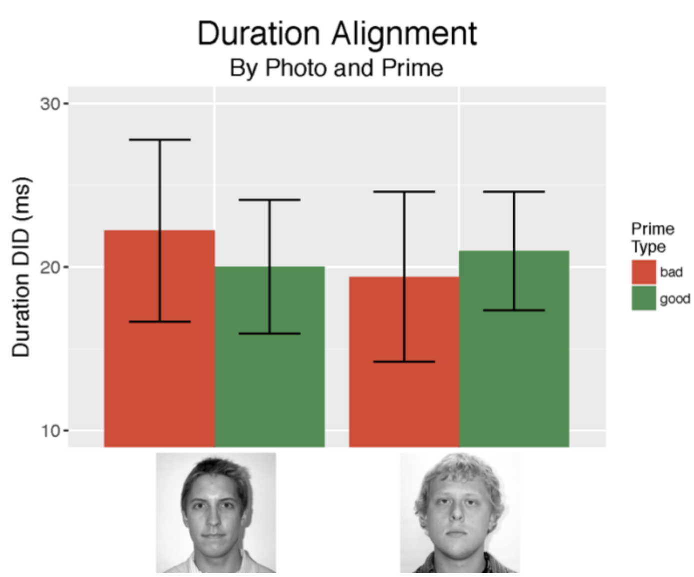

Figure 16. Interaction between photo and prime type on vocal alignment to duration.

interaction was not significant, β=3.52, *t*=.34, *p*=.733, and neither the main effects nor interaction terms improved MEM fit, both χ^2<.28, p>.59 (Figure 16). This indicates that the photo selection criteria were successful because the particular face that was paired with a prime did not differentially influence alignment behavior. Whatever impressions may have been driven independently by the facial features were equated across both photos.

Individual Differences

Baseline variability.

As with Experiment 1, an analysis of the extent to which baseline variation in duration could account for alignment to duration was conducted, with the prediction that more variation would be related to greater alignment. Once again, variation was operationalized as standard deviation (SD) of baseline trials (Figure 17). No statistically significant association between baseline variability in duration and alignment to duration was found, β=-.01, *t*=-.30, *p*=.763. Including SD did not improve MEM fit, $\chi^2(1)$=.09, *p*=.762. The effect remained non-significant even when 4 apparent outliers (Duration SD > 400 ms) were excluded, β=.02, *t*=.27, *p*=.79. This

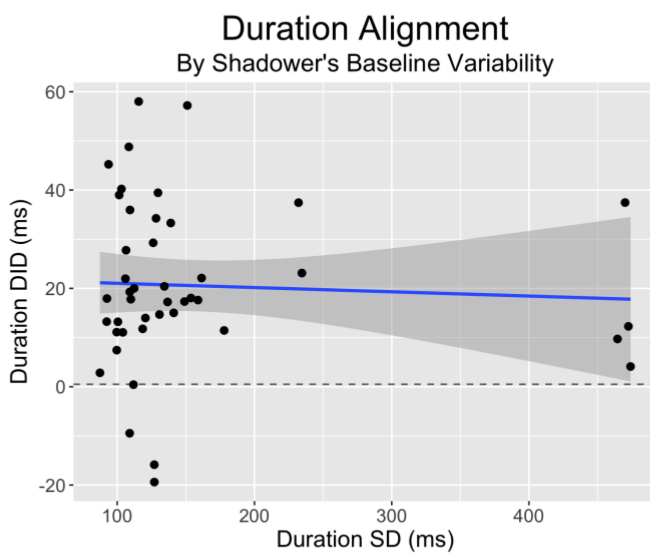

Figure 17. Relationship between baseline variation (duration standard deviation) and vocal alignment to duration (duration DID). Unless otherwise noted, the blue line indicates the best fit regression line and grey band indicates the 95% confidence interval.

suggests that there was not a relationship between variability and alignment to duration as indexed by SD and DID, respectively.

Empathy Quotient.

To address the question of whether or not individual differences in empathy affect the propensity to engage in vocal alignment, the data from Experiment 1, 2, and the preliminary experiment were pooled for a single analysis. Aggregating the data permitted analysis of dispositional empathy and vocal alignment in 99 shadowers. Only the relationship between Empathy Quotient (EQ) and duration DID was considered because it was measured in both main experiments and the preliminary experiment[6]. As with previous analyses, the data were analyzed with GLMM and maximal random effects structure justified by the sample was used[7].

Mean score on the EQ was 45.45 (median 47), with EQ ranging between 13 and 71. The relationship between EQ, cognitive load, and duration alignment is

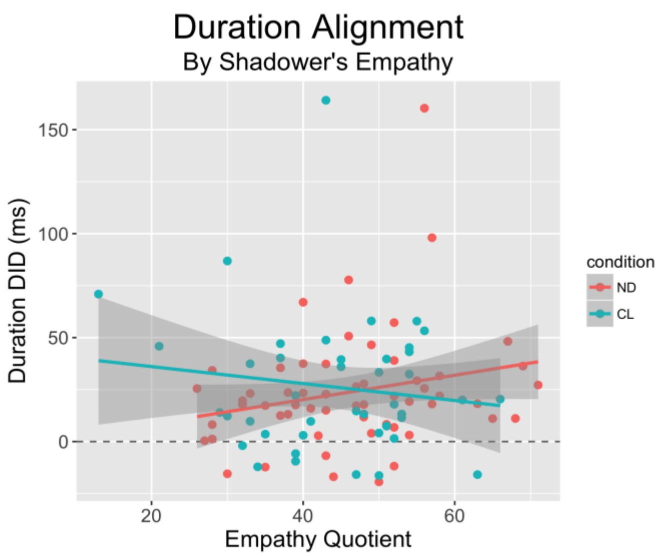

Figure 18. Relationship between shadower's empathy score (Empathy Quotient), load condition (CL or ND), and vocal alignment to duration (duration DID).

---

[6] The reader should note that duration was not the only assay of alignment in Experiment 1. As such, naturalistic variation in model talker durations were used in Experiment 1, whereas duration was artificially varied in Experiment 2 and in the preliminary experiment.

[7] RE structure for this analysis was (1 | shadower) + (1 | word) + (1 | model talker) + (1 | experiment) + (1 | cognitive load condition). Random intercepts for model talker within shadower, or shadower within experiment produced convergence errors. The cognitive load condition for all subjects in the preliminary experiment was coded as ND, as all participants completed the task without load.

depicted in Figure 18. The visual trend shows differing linear relationships between EQ and alignment to duration across load conditions. Analysis revealed no main effect of empathy, β_{EQ}=.10, *t*=.365, *p*=.716. EQ also did not improve MEM fit, $\chi^2(1)$=.38, *p*=.5396. The main effect of cognitive load was marginal in this analysis, β_{CL}=45.82, *t*=1.87, *p*=.064; $\chi^2(1)$=.35, *p*=.55, and this main effect was conditioned by a marginal empathy by load interaction, β_{CLxEQ}=-.93, *t*=-1.76, *p*=.08. The MEM containing the interaction term marginally outperformed one that excluded it, $\chi^2(1)$=3.07, *p*=.08.

The data were divided by load condition for follow-up analysis. For the CL condition, the effect of empathy was non-significant both within the MEM (β_{EQ}=-.37, *t*=-.83, *p*=.409) and in MEM comparisons ($\chi^2(1)$=.69, *p*=.41). For the ND condition, there was a marginal, positive relationship between empathy and duration alignment, β_{EQ}=.56, *t*=1.78, *p*=.081. EQ marginally improved MEM fit, $\chi^2(1)$=3.08, *p*=.079. To summarize, there was no overall indication that higher empathy was associated with greater vocal alignment, at least for the indices of empathy and alignment assessed here. However, the effect of empathy may be conditioned by cognitive load.

Discussion

A wealth of literature supports the hypothesis that macro-level social categories and stereotypes affect vocal alignment, as does liking (Abrego-Collier et al., 2011; Babel, 2012; Babel et al., 2014; Bourhis & Giles, 1977; Kim, 2012; Namy et al., 2002; Sidaras, 2011; Yu et al., 2013). Experiment 2 specifically examined whether positive or negative secondhand experiences would influence the degree to which shadowers align. In this experiment, social information obtained from faces and voices was held constant while individual social characteristics or “character” was manipulated. Furthermore, this experiment evaluated whether alignment based on implied traits required cognitive resources. A sample of subjects participated in either a single-task or a dual-task paradigm that introduced a cognitive load and introduced positive and negative information about each model talker. This design permitted an examination of (1) whether talker evaluations influence the likelihood that language users will entrain to speaking rate and (2) whether using trait inferences to guide alignment behavior is efficient.

Secondhand Experience of Others

Although previous work suggests that the introduction of social evaluation may influence patterns of alignment (Abrego-Collier et al., 2011, Babel, 2012; Babel et al., 2010; Yu et al., 2013), most studies have defined social information by categories or stereotypes of groups of speakers, rather than specific attributes of the individual speaker. Experiment 2 replicated and extended existing work on vocal alignment by examining the influence on alignment of trait inferences made from short, diagnostic narratives. First, the findings suggest that shadowers inferred traits from the narratives. Shadowers who reported a preference significantly preferred the talker who was associated with positive trait inferences. Second, these inferences appeared to influence alignment behavior. In the no load condition, shadowers showed greater vocal alignment to the good talker than to the bad talker. The findings demonstrate that trait inferences made from minimal, secondhand input may act as an important constraint on the extent to which interlocutors engage in vocal alignment. This suggests that language users can adjust alignment behavior based on secondhand, social information not necessarily obtained by activating social stereotypes.

Experiment 2 also offers some insight into whether or not social evaluations require attentional or cognitive resources to affect speech alignment. While participants in both conditions were able to form preferences, shadowers under cognitive load more frequently misidentified which talker had been presented as good and bad. This suggests both that trait inferences did influence preferences and that talker information was harder to track when processing demands on participants were high. This implies at least two possibilities for the manner in which social information affects vocal alignment. Although all participants made inferences without load present, it is possible that subsequent cognitive load inhibited rehearsal, maintenance, or retention of associations between a model talker and their inferred trait. While some evidence suggests that social stereotypes are automatically activated (Bargh, 1994; Chartrand & Bargh, 1999) and can influence behavioral mimicry (Bargh et al., 1996), trait inferences like the kind examined in Experiment 2 may decrease under high processing demands (Crawford, Skowronski, Stiff, & Scherer, 2007; Ramos et al., 2012; Wells et al., 2011; Uleman et al., 1992). The finding suggests that using certain kinds of social information to guide alignment behavior may require cognitive resources. These results could be interpreted as evidence for social information acting a secondary process modulating an automatic vocal alignment behavior. Alternatively, participants in the cognitive load condition may have been able to make trait inferences and map the inferred traits to the appropriate talkers, but were not able to use the information to alter their behavior.

The current experiment cannot adjudicate between whether it is the social information itself or another process that links social information to behavioral adjustment. However, the data nonetheless support the proposal that vocal alignment is modulated by social information in some way. Additional evidence for this interpretation can be found in the interaction that was observed between prime type and cognitive load on duration DID. Without load, shadowers aligned more to the good talker's duration, but under load, there was no reliable difference between alignment to the good and bad talkers. The absence of an effect suggests that participants in the ND condition were using social information to modulate their alignment behavior. When processing resources were less available, the alignment difference between good and bad talkers disappeared, suggesting that participants were no longer using social information to differentially align to the good and bad talkers.

Individual Differences

Previous evidence shows that individuals differ in their proclivity to vocally align (Pardo, 2013; Pardo et al., 2017). Some of this variability may be explained by dispositional differences that affect one's baseline tendency to align (Aguilar et al., 2015; Yu et al., 2013) or by differences in speech production habits that affect motoric ability to align (Babel, 2010; Kim et al., 2011; Walker & Campbell-Kibler, 2015). The current experiment tested individual differences in speech production and dispositional empathy and their relationship to vocal alignment.

First, a shadower's baseline variability in speech production and their alignment to the specific speaking rates of the model talkers (duration DID) was tested. No relationship was found between individual variability in speech production on vocal alignment. The absence of an effect in the current investigation perhaps reflects the different operationalization of 'distance from a target' and 'habitual production repertoires' than was adopted by Walker and Campbell-Kibler (2015). Those authors defined 'distance' across two pairs of English dialects: U.S. Inland North and U.S. Midland, New Zealand and Australia. 'Habitual production repertoires' were loosely defined as the shadower's vowel spectra based on dialect region (New Zealand or U.S. Midland). The current investigation took somewhat narrower definitions of both constructs, in as much as 'distance' referred to baseline distance between model and shadower and 'variability' referred to habitual speaking rate variation. When defined in this way, variability, at least along this dimension, may not be predictive of individual tendency to align.

Next, data were pooled across experiments to test the influence of individual differences in empathy. There was no overall relationship between differences in empathy and alignment on speaking rate. This finding is somewhat surprising given that the literature suggests a connection between empathy and general behavioral alignment (de Guzman et al., 2016; Kühn et al., 2010; Meltzof & Decety, 2003; Stiff et al., 198l; van Baaren et al., 2003). However, it is possible that the manner in which empathy influences vocal alignment may be attentionally or cognitively demanding. Indeed, there was a marginally significant interaction between cognitive load condition and empathy, such that empathy only influenced alignment behavior for shadowers in the no load condition. This result implies that dispositional differences in empathy may not affect baseline rate of alignment, but rather may serve as an additional type of cognitive modulation. Such a proposal could also explain why there was no overall effect of empathy on vocal alignment, as 42 of the 99 participants included in the empathy analysis completed the alignment task under cognitive load. The necessary processing resources for more effortful components of empathy may have been disrupted by load. It is also possible that global empathic tendencies as measured by the Empathy Quotient (EQ) are not strongly related to vocal alignment, but other capacities related to theory of mind and empathy may potentially be.

Evidence for this possibility comes from work on behavioral mimicry. In a 1999 study, participants completed a picture description task with a confederate partner, and the participant's mimicry of the confederate was assessed (Chartrand & Bargh, 1999). The participant also completed the Interpersonal Reactivity Index (IRI; Davis, 1980). The IRI does not measure empathy per se, but measures related constructs on subscales that tap empathic concern, perspective-taking, personal distress, and fantasy. Chartrand and Bargh (1999) found that the perspective-taking subscale, but not empathic concern, was associated with participants' tendency to engage in behavioral alignment. The EQ was chosen specifically because it measures a combination of cognitive and emotional empathic tendencies but not other, related constructs (Baron-Cohen & Wheelwright, 2004). It is, however, correlated with both the perspective-taking ($r=.485$; Lawrence, Shaw, Baker, Baron-Cohen, & David, 2004) and empathic concern ($r=.423$, Lawrence et al., 2004) subscales of the IRI.

Summary

The results of Experiment 2 suggest that (1) information about a specific individual gained from trait inferences rather than general social categories can influence patterns of vocal alignment, (2) adjusting alignment to track social information may require processing resources and serves to modulate vocal alignment under normal processing demands, and (3) individual differences in vocal flexibility or in empathic tendencies did not predict vocal alignment in this minimally social paradigm. This experiment demonstrates a modulatory role for social information in what may otherwise be an automatic alignment process.

General Discussion

Across a variety of social species, vocalizations can be used to distinguish groups from one another and entrainment to vocal characteristics may influence social bonds at the individual level. In humans, the mechanisms that support this important communicative function have traditionally been understood at two different levels, which either emphasize the socio-motivational features of alignment or highlight the general principles of cognition such as perceptual-motor links that may support alignment. Rather than examine these constructs separately, this dissertation offered a unified framework for understanding the mechanisms underlying vocal alignment.

Drawing on parallels between alignment to vocal and non-vocal behavior demonstrated in the literature, I proposed that vocal alignment may be understood as an automatic, non-resource demanding process modulated by secondary, goal-related features. This framework is supported by two experiments that explored the extent to which general alignment processes and the use of social information during alignment changed under cognitive load. Based on the results of these two experiments, I conclude that vocal alignment may be a relatively effortless, non-resource demanding perceptual-motor process that is subject to modulation by social information under typical cognitive processing demands.

Automaticity of Vocal Alignment

Researchers have argued that the construct of “automaticity” can be broken down into subparts, and rather than describe processes as “automatic” in an all-or-none fashion, processes can be more appropriately understood in terms of degrees of automaticity

(Bargh, 1994; Moors & De Houwer, 2006). Past studies of vocal alignment provide evidence that alignment behaviors can occur without *awareness* (Garnier et al., 2013; Goldinger, 1998) or *intent* (Delvaux & Soquet, 2007; Evans & Iverson, 2007; Pardo et al., 2012) – two criteria of automaticity. However, the extent to which vocal alignment is *efficient* (can occur without the use of effortful or attentional resources) has received mixed evidence in the literature. Two previous studies have examined this question.

Abel and Babel (2017) found that increased difficulty of a dyadic Lego construction task in which one partner explained the steps for model construction to the other led to less vocal alignment between task partners. That is, the most vocal alignment was observed in the easiest condition. By contrast, Solanki et al. (2016) found that greater difficulty in a picture comparison task was associated with increased alignment. In both of these cases, the difficulty of the task was conditioned to some extent on the dyads themselves, and processing demands may not have been equally high for both partners. Especially in Abel and Babel's (2017) study, the partner charged with describing the model construction may have been under much greater processing load than the partner whose job was to follow instructions. The current experiments contribute to the existing literature on the efficiency of vocal alignment by manipulating cognitive load independently of the vocal alignment task. In Experiment 1, it was shown that listener perceptions of alignment did not significantly differ by cognitive load condition, suggesting that alignment may be an efficient process by Bargh's (1994) definition.

Although the effect of cognitive load did not influence perceived patterns of alignment in Experiment 1, Experiment 2 demonstrated that cognitive load impaired differential alignment to "good" and "bad" talkers on duration. Whereas shadowers aligned more to the good talker when processing demands were low, shadowers aligned equivalently to the good and bad talkers under cognitive load. This was the case even though the initial presentation and encoding of social information occurred when participants were not engaged in the demanding secondary task. Because shadowers obtained social information without a load present, the effect of cognitive load on the subsequent use of social information in vocal alignment suggests two possibilities. On the one hand, the use of social information to guide alignment behavior may be resource demanding, which would imply that social information itself serves as a modulator of vocal alignment. On the other hand, cognitive load may make it more difficult for shadowers to remember the social characteristics ascribed to each talker, suggesting that more general working memory or attentional mechanisms modulate the interface between social information and vocal alignment. Social priming accounts suggest that social judgment and biases may act directly on behavior (Chartrand & Bargh, 1999; Chen & Bargh, 1997), but it is possible that the decision-based processes associated with trait inference may require attentional resources to affect alignment. Although both cases are consistent with the proposal that the automatic tendency to align is modulated by a secondary process, disambiguating between the maintenance of social information in memory and the use of social information to direct vocal alignment will be important for understanding mechanism.

A noteworthy assumption of the current investigation is that cognitive load as instantiated by a visuospatial working memory task should influence speech alignment. Implicitly, the approach taken here is consistent with a shared resources model of processing, where attentionally demanding tasks rely on a shared pool of resources

(Hiraga, Garry, & Carson, & Summers, 2009; Tombu & Jolicoeur, 2003; Wickens, 1991). This assumption is not without empirical basis: for example, in an investigation of automaticity of behavioral imitation, van Leeuwen and colleagues (2009) used a verbal N-back task (matching an aurally presented letter to two or three letters ago) as the memory load for a finger-tapping (visuospatial imitation) task. Others have found that visually presented distractors (e.g., visual search task) impaired the ability to detect an aurally presented tone (e.g., auditory detection; Raveh & Lavie, 2015). However, a key difference between Raveh and Lavie (2015)'s study and the present investigation is that rather than targeting working memory, their visual search task may have more directly impacted attention. It is possible that other kinds of or timing of load may have differential effects on vocal alignment. For example, participants in Mattys, Barden, and Samuel's (2014) study completed a visual search task *concurrent* with the presentation of an auditory stimulus. In their experiment, the ability to distinguish fine-grained acoustic features decreased as demands of the visual search task increased. Extending the current findings to one or more kinds of cognitive load would strengthen the claim that the absence of main effects of load effect in Experiments 1 and 2 was due specifically to a non-resource demanding alignment process.

Individual Differences

Given the individual variation in the tendency to vocally align, researchers have sought out intrinsic individual traits that contribute to vocal alignment behavior. Dispositional characteristics and personality traits have received some attention (Yu et al., 2013). For example, in a dyadic task setting, Kurzius (2015) showed that extraversion and openness to experience both significantly predicted alignment to speaking rate of confederates. In addition to effects of certain personality dimensions, dispositional sensitivity to approval from others has been shown to influence degree of alignment (Aguilar et al., 2016). This dissertation sought to contribute to the study of individual differences by examining empathic feeling and thinking. Using the Empathy Quotient (EQ; Baron-Cohen & Wheelwright, 2004), the findings provided minimal support for a global role of empathy in vocal alignment. This was unexpected, given past evidence suggesting that individuals unintentionally shadow subtle emotional prosody in speech (Neuman & Strack, 1998) and that perspective-taking predicts extent of behavioral imitation (Chartrand & Bargh, 1999). However, the current study uncovered a relationship between empathy and vocal alignment that was conditioned by cognitive load, indicating that empathic feeling and thinking may require cognitive resources to guide alignment behavior. The results suggest that empathy may affect vocal alignment behavior, but does not necessarily affect the base tendency to align (e.g., a person with high empathy will not always align more, regardless of other demands).

With respect to individual differences in motoric or perceptual abilities, there have been suggestions in the literature that the differences in speech ability may be tied in some way to propensity to engage in vocal alignment. The proposal seems intuitive; vocal alignment, a speech production behavior, should be influenced by individual differences in speech production. Past research has found some support for this, in that individuals higher on "language talent" may vocally align to a greater extent than their less talented counterparts (Lewandowski, 2009). Part of this effect may be accounted for by *linguistic distance*, or how much distinctiveness or overlap is present in the sound repertoires of individuals with differing their language backgrounds (Kim, 2011; Walker & Campbell-Kibler, 2015). For example, Walker and Campbell-Kibler (2015) demonstrated that individuals who share a native language but have different dialects (e.g., US Midland versus New Zealand) may align more than individuals with two very similar dialects (e.g., US Midland versus US Inland North). However, distance may not be as important as range; the ability to hit certain acoustic-phonetic targets may account for the effect of "language talent" on alignment (Lewandowski & Nygaard, under review). The current experiments examined this proposal, but found only weak supporting evidence. Although positive associations were generally noticed between baseline acoustic variability and vocal alignment, these trends did not achieve statistical significance. Discussed below, another individual difference likely to be important for vocal alignment is the ability to detect fine-grained characteristics of the voice.

Speech Perception and Production

Vocal alignment represents one of many behaviors in the speech perception and production repertoire. Considered in the context of speech more broadly, the current results support the importance of sensitivity to talker characteristics for speech perception, but also demonstrate that talker traits have consequences for subsequent speech production behavior. Past research has demonstrated that listeners are sensitive to the acoustic properties associated with talkers' voices (Bradlow, Nygaard, & Pisoni, 1999; Creel & Bregman, 2011; Goldinger, Pisoni, & Logan, 1991; Nygaard, Sommers, & Pisoni, 1994; Nygaard, Burt, & Queen, 2000). For example, listeners remember items in lists better if they are spoken by a single talker than if the items are spoken by multiple talkers (Martin, Mullennix, Pisoni, & Summers, 1989), are faster to recognize old words when they are repeated in the original voice (Palmeri, Goldinger, & Pisoni, 1993), and are able to use talker identity as a cue for predicting word identity (Creel, Aslin, & Tanenhaus, 2008). Past findings support the notion that features of talker's voice are encoded during the perception of speech. Vocal alignment shows that the characteristics of talker's voice are not only encoded, but consequently influence speech production.

In Experiment 1, vocal alignment to native- and Spanish-accented model talkers was examined. Shadowers aligned marginally more to the Spanish-accented speakers than to native-accented speakers. The acoustic features of Spanish-accented model talkers were more different from the native-accented shadowers, and independent raters confirmed that the accented models did indeed sound more accented than the native models. Listeners were sensitive to these differences, suggesting that the atypical characteristics of the Spanish-models' voices may have been encoded and resulted in subsequent alteration of speech alignment (Babel et al., 2014; Bradlow et al., 1999). The results of Experiment 1 in particular point to the importance of the shared or linked architecture of speech perception and production as a mechanism for alignment behavior (Goldinger & Azuma, 2004; Fowler & Galantucci, 2005), though it does not rule out the role of general priming mechanisms across levels of language (Pickering & Garrod, 2004; 2006; 2013).

The role of low-level principles of cognition highlight the importance of perception to fine-grained speech structure in determining vocal alignment. Recent evidence has suggested that cognitive load impairs attention to fine-grained acoustic details (Mattys & Palmer, 2015; Mattys & Wiget, 2011). Nonetheless, shadowers experiencing cognitive load in Experiments 1 and 2 showed vocal alignment on speaking rate and pitch. In Experiment 1, cognitive load increased vocal alignment on pitch and duration, and in Experiment 2, cognitive load did not decrease vocal alignment to duration. If it is indeed the case that listeners under cognitive load have less access to details about the sound structure of words, then the current study suggests that conscious awareness of the acoustics of speech may not be necessary in order to vocally align. Research in the domain of perceptual learning has revealed similar findings; that is, explicit attention to speech may not be necessary to gain the benefits of exposure to a novel accent (Seitz et al., 2010). The current findings further suggest that attention may not be necessary to translate perceived speaking styles into speech production, pointing to a tight relationship between speech perception and speech production more generally.

Speech Specificity

A question not directly addressed by the current study pertains to whether vocal alignment is part of a general suite of behavioral mimicry behaviors. Low-level priming and perceptual-motor couplings are not exclusive to speech, and as such, may be the shared set of mechanisms underlying various kinds of alignment behaviors. A handful of studies have shown that alignment to speech characteristics, such as lexical choice, and alignment to non-verbal behaviors, such as manual gesture and postural sway, occur simultaneously (Louwerse, Dale, Bard, & Jeuniaux, 2012; Shockley, Baker, Richardson, & Fowler, 2007). Although alignment to speech and non-verbal behavior were not directly compared in the current study, conventional wisdom may offer some insight into this question: if it looks like a duck and quacks like a duck, then it probably is a duck.

That is, if similar factors cause similar patterns of alignment for speech and non-speech behavior, this may be support for similar underlying mechanisms.

The current dissertation suggests that vocal alignment shares some similarities with alignment to non-verbal behavior. For example, non-verbal imitation (e.g., finger-tapping) has been shown to increase under a high cognitive load (van Leeuwen et al., 2009; Catmur, 2016). Experiment 1 demonstrated that vocal alignment to speaking rate and pitch (though not perceived alignment) increased under cognitive load. Furthermore, Experiment 2 demonstrated an effect of social information on alignment, consistent with patterns of alignment to non-verbal behavior (Chartrand & Bargh, 1999; Shockley et al., 2007). Findings from the current experiments provide a foundation for future work that may compare vocal alignment to other types of behavioral alignment. I speculate that alignment tendencies may be common across perceptual-motor domains.

However, even if alignment is shared across perceptual-motor systems and similar mechanisms and constraints act on all entrainment behavior, aspects of vocal alignment may yet be unique. Alignment to speech may have broader consequences not achieved by non-verbal alignment by virtue of its effects on linguistic and social communication. Because linguistic features carry social and group meaning, vocal alignment might have larger consequences for individual (Cutler, 2010; Lewandowski & Nygaard, under review; Pardo et al., 2012) and group identity (Giles et al., 1991; Bourhis & Giles, 1973; Evans & Iverson, 2007; Delavoux & Soquet, 2007; Sancier & Fowler, 1997). Over time, the speech perception and speech production system may contribute to historical language change (Trudgill, 2008; Tuten, 2008). Although general behavioral imitation may improve feelings of perceived closeness and rapport across individuals, behavioral imitation may not shape the social communicative landscape to the same extent as alignment to vocal and speech cues.

Limitations

A key limitation of the current experiments pertains to the number of model talkers used. Although consistent with the numbers of model talkers used in comparable studies (Abrego-Collier et al., 2011, Brouwer et al., 2010; Babel, 2010; 2012; Kim, 2011; Namy et al., 2002), it is a challenge to draw generalizable conclusions about patterns of alignment from small subsets of talkers. The consequences of this limitation may be somewhat attenuated by the goal of the current experiments. Experiment 2, in particular, was designed to assess alignment based on individual rather than group characteristics. However, future work must ensure that patterns of alignment are consistent across similar trait inferences and similar groups of talkers.

Another limitation concerns ecological validity. The design of both experiments was structured to maximize experimental control over the factors of interest, but as a consequence, the paradigm was not naturalistic. In a socially minimal setting such as the shadowing task paradigm, the overall effect of explicit social information may be diminished (Lewandowski & Nygaard, under review). For example, as mentioned previously, in a live interview setting where native Welsh-speaking adults “overheard” (this was a staged event) a high-status interviewer insulting Welsh language and culture, Welsh speakers who highly valued their cultural identity vocally diverged from the interviewer. In a more comparable paradigm to the current experiment, Babel (2010) primed participants with a positive or negative stereotype of Australians prior to a naming task. Babel found that enduring attitudes as measured by a post-task IAT were better predictors of alignment to an Australian talker than the primes. The effect of the character information may have had a stronger effect had the experience been presented in a more naturalistic, interpersonal way. Such an approach was not pursued in this set of studies because elements of the experimenter and participant's interaction (e.g., level of trust, liking, etc.) may have introduced unknown variables.

Conclusion

The current work offers an approach to the study of vocal alignment that highlights the interplay of cognitive, perceptual, and social mechanisms underlying entrainment in spoken language. By framing the question of mechanism as one of gradations of automatic processing, this dissertation explores the automaticity of both social and perceptual-motor processes related to vocal alignment. Taking inspiration from work on non-speech imitation, the current dissertation provides evidence for an automatic, perceptual-motor component and an effortful, socio-motivational component to alignment behavior. Entrainment to others' voices serves a variety of social, linguistic, and instructive functions across species, and it is crucial to understand the underlying mechanisms. These experiments suggest that vocal alignment is based on the interaction of general cognitive mechanisms and goal-related functions. Understanding these mechanisms may improve the efficacy of communication in business, public health, and educational settings.

References

Abel, J., & Babel, M. (2017). Cognitive load reduces perceived linguistic convergence between dyads. *Language and Speech*, *60*(3), 479-502.

Abrego-Collier, C., Grove, J., Sonderegger, M., & Yu, A. C. (2011). Effects of speaker evaluation on phonetic convergence. In *Proceedings of the 17th International Congress of the Phonetic Sciences* (pp. 192-195).

Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation improves language comprehension. *Psychological Science*, *21*(12), 1903-1909.

Aguilar, L., Downey, G., Krauss, R., Pardo, J. S., Lane, S., & Bolger, N. (2016). A dyadic perspective on speech accommodation and social connection: both partners' rejection sensitivity matters. *Journal of Personality*, *84*, 165-177.

Alais, D., Morrone, C., & Burr, D. (2006). Separate attentional resources for vision and audition. *Proceedings of the Royal Society of London B: Biological Sciences*, *273*(1592), 1339-1345.

Allport, G. W. (1954). *The nature of prejudice.* Oxford, England: Addison-Wesley.

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., & Weinert, R. (1991). The HCRC map task corpus. *Language and Speech*, *34*(4), 351-366.

Anderson, E., Siegel, E. H., Bliss-Moreau, E., & Feldman-Barrett, L. (2011). The visual impact of gossip. *Science*, *332*, 1446-1448.

Baayen, R. H. (2011). languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics". R package version 1.4. (http://CRAN.R-project.org/package=languageR).

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412.

Baayen, R. H., & Milin, P. (2015). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12-28.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*(1), 177-189.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, *39*, 437-456.

Babel, M., & Bulatov, D. (2012). The role of fundamental frequency in phonetic accommodation. *Language and Speech*, *55*(2), 231-248.

Babel, M., McGuire, G., Walters, S., & Nicholls, A. (2014). Novelty and social preference in phonetic accommodation. *Laboratory Phonology*, *5*(1), 123-150.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445-459.

Balsby, T. J., & Adams, D. M. (2011). Vocal similarity and familiarity determine response to potential flockmates in orange-fronted conures (Psittacidae). *Animal Behaviour*, *81*(5), 983-991.

Balsby, T. J., & Scarl, J. C. (2008). Sex-specific responses to vocal convergence and divergence of contact calls in orange-fronted conures (Aratinga canicularis). *Proceedings of the Royal Society of London B: Biological Sciences, 275*(1647), 2147-2154.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition volume 1: Basic processes* (2nd ed.), 1-40. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bargh, J. A. (1982). Attention and automaticity in the processing of self-relevant information. *Journal of Personality and Social Psychology*, *43*(3), 425.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*(2), 230-244.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163-175.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high functioning autism, males, females, scientists and mathematicians. *Journal of Autism & Developmental Disorders*, *31*, 5-17.

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

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-4. (https://github.com/lme4/lme4/).

Beal, B. (2018). Investigation of character detection. Unpublished Study.

Bilous, F. R., & Krauss, R. M. (1988). Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads. *Language and Communication, 8*(3), 183-194.

Bliss-Moreau, E., Barrett, L. F., & Wright, C. I. (2008). Individual differences in learning the affective value of others under minimal conditions. *Emotion*, 8(4), 479-493.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*, 355-387.

Boersma, P., & Weenink, D. (2014). Praat: Doing phonetics by computer [Computer program]. Version 5.4.18, retrieved 3 February 2014 from http://www.praat.org/.

Bourhis, R. Y., & Giles, H. (1977). The language of intergroup distinctiveness. *Language, Ethnicity and Intergroup Relations*, *13*, 119-135.

Bradbury, J. W., Cortopassi, K. A., & Clemmons, J. R. (2001). Geographical variation in the contact calls of orange-fronted parakeets. *The Auk*, *118*(4), 958-972.

Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America*, *97*(3), 1916-1924.

Bradlow, A. R., Nygaard, L. C., & Pisoni, D. B. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics*, *61*(2), 206-219.

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, *20*(3), 255-272.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75*(2), B13-B25.

Branigan, H. P., Pickering, M. J., Pearson, J., McLean, J. F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition*, *121*(1), 41-57.

Brouwer, S., Mitterer, H., & Huettig, F. (2010). Shadowing reduced speech and alignment. *The Journal of the Acoustical Society of America*, *128*(1), EL32-EL37.

Burgess, C., & Livesay, K. (1998). The effect of corpus size in predicting reaction time in a basic word recognition task: Moving on from Kučera and Francis. *Behavior Research Methods, Instruments, & Computers, 30*(2), 272-277.

Campbell, M. W., & De Waal, F. B. (2011). Ingroup-outgroup bias in contagious yawning by chimpanzees supports link to empathy. *PloS one*, *6*(4), e18283.

Carlston, D. E., & Skowronski, J. J. (2005). Linking versus thinking: Evidence for the different associative and attributional bases of spontaneous trait transference and spontaneous trait inference. *Journal of Personality and Social Psychology*, *89*(6), 884-898.

Carlston, D. E., & Skowronski, J. J. (1994). Savings in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology, 66*(5), 840-856.

Catmur, C. (2016). Automatic imitation? Imitative compatibility affects responses at high perceptual load. *Journal of Experimental Psychology: Human Perception and Performance, 42*(4), 530-539.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: the perception-behavior link and social interaction. *Journal of Personality and Social Psychology, 76*(6), 893-910.

Chartrand, T. L., & Lakin, J. L. (2013). The antecedents and consequences of human behavioral mimicry. *Annual Review of Psychology*, *64*, 285-308.

Chartrand, T. L., Cheng, C. M., Dalton, A. N., & Tesser, A. (2010). Nonconsious goal pursuit: Isolated incidents or adaptive self-regulatory tool?. *Social Cognition*, *28*(5), 596-588.

Chartrand, T. L., Maddux, W. W., & Lakin, J. L. (2005). Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry. In R. R. Hassin, J. S. Uleman, and J. A. Bargh (Eds.), *The new unconscious* (pp. 334-361). New York, NY: Oxford University Press.

Chen, M., & Bargh, J. A. (1997). Nonconscious behavioral confirmation processes: The self-fulfilling consequences of automatic stereotype activation. *Journal of Experimental Social Psychology*, *33*(5), 541-560.

Coupland, N. (1984). Accommodation at work: Some phonological data and their implications. *International Journal of the Sociology of Language*, *1984*(46), 49-70.

Crawford, M. T., Skowronski, J. J., Stiff, C., & Scherer, C. R. (2007). Interfering with inferential, but not associative, processes underlying spontaneous trait inference. *Personality and Social Psychology Bulletin*, *33*(5), 677-690.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, *106*(2), 633-664.

Creel, S. C., & Bregman, M. R. (2011). How talker identity relates to language processing. *Language and Linguistics Compass*, *5*(5), 190-204.

Crockford, C., Herbinger, I., Vigilant, L., & Boesch, C. (2004). Wild chimpanzees produce group-specific calls: a case for vocal learning?. *Ethology*, *110*(3), 221-243.

Cutler, C. (2010). Hip-hop, white immigrant youth, and African American Vernacular English: Accommodation as an identity choice. *Journal of English Linguistics*, *38*(3), 248-269.

Davis, M. H. (1994). *Empathy: A social psychological approach.* Madison, WI: Westview Press.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology, 10*, 85.

Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica*, *64*(2-3), 145-173.

Denes P., & Pinson E. (1963). *The speech chain.* Garden City: Anchor Press/Doubleday.

DeSteno, D., Breazeal, C., Frank, R. H., Pizarro, D., Baumann, J., Dickens, L., & Lee, J. J. (2012). Detecting the trustworthiness of novel partners in economic exchange. *Psychological Science, 23*(12), 1549-1556.

Devine, P. G. (1989). Stereotypes and Prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*(1), 5-18.

de la Torre, S., & Snowdon, C. T. (2009). Dialects in pygmy marmosets? Population variation in call structure. *American Journal of Primatology*, *71*(4), 333-342.

De Guzman, M., Bird, G., Banissy, M. J., Catmur, C. (2016). Self-other control processes in social cognition: from imitation to empathy. *Philosophical Transactions of the Royal Society B, 371*, 1-9.

De Waal, F. B. (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology*, *59*, 279-300.

Dijksterhuis, A., & van Knippenberg, A. (1998). The relation between perception and behavior, or how to win a game of Trivial Pursuit. *Journal of Personality and Social Psychology, 74*, 865-877.

Doyen, S., Klein, O., Pichon, C. L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind?. *PloS one, 7*(1), e29081.

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.

Duff, K. J., & Newman, L. S. (1997). Individual differences in the spontaneous construal of behavior: Idiocentrism and the automatization of the trait inference process. *Social Cognition*, *15*(3), 217-241.

Eisenberg, N., & Miller, P. A. (1987). The relation of empathy to prosocial and related behaviors. *Psychological Bulletin, 101*(1), 91-119.

Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: a study of accent change in young adults. *Journal of the Acoustical Society of America, 121*(6), 3814-3826.

Fischer, J., Wheeler, B. C., & Higham, J. P. (2015). Is there any evidence for vocal learning in chimpanzee food calls? *Current Biology*, *25*, R1028-R1029.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America, 106*(3), 1511-1522.

Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language*, *49*(3), 396-413.

Fowler, C. A., & Galantucci, B. (2005). The relation of speech perception and speech production. *The Handbook of Speech Perception*, 632-652.

Garnier, M., Lamalle, L., & Sato, M. (2013). Neural correlates of phonetic convergence and speech imitation. *Frontiers in Psychology*, *4*(600), 1-15.

Gawronski, B., & Bodenhausen, G. V. (2014). Implicit and explicit evaluation: A brief review of the associative-propositional evaluation model. *Social and Personality Psychology Compass*, *8*(8), 448-462.

Gazzaniga, M. S., Ivry, B., & Mangun, G. R. (2009). *Cognitive neuroscience: The biology of the mind (2nd ed.)*. New York, NY: W. W. Norton & Company.

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology, 60*(4), 509-517.

Giles, H. (1973). Accent mobility: a model and some data. *Anthropological Linguistics, 15*(2), 87-105.

Giles, H., Coupland, N., & Coupland, J. (1991). *Contexts of accommodation: Developments in applied sociolinguistics*. Cambridge, UK: Cambridge University Press.

Giles, H., & Ogay, T. (2007). Communication accommodation theory. In B. B. Whaley & W. Samter (Eds.), *Explaining communication: Contemporary theories and exemplars*, Mahwah, NJ: Psychology Press.

Giles, H., Willemyns, M., Gallois, C., & Anderson, M. C. (2007). Accommodating a new frontier: The context of law enforcement. In K. Fiedler (Ed.), *Social Communication* (pp. 129-162). New York, NY: Psychology Press.

Gluszek, A., & Dovidio, J. F. (2010). 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.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251-279.

Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin & Review, 11*(4), 716-722.

Goldinger, S. D., Pisoni, D. B., & Logan, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(1), 152-162.

Gravano, A., Levitan, R., Willson, L., Beňuš, Š., Hirschberg, J., & Nenkova, A. (2011). Acoustic and prosodic correlates of social behavior. In *Twelfth Annual Conference of the International Speech Communication Association, 2011,* 97-100.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4-27.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464-1480.

Gregory Jr, S. W., Dagan, K., & Webster, S. (1997). Evaluating the relation of vocal accommodation in conversation partners' fundamental frequencies to perceptions of communication quality. *Journal of Nonverbal Behavior*, *21*(1), 23-43.

Gregory Jr, D., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status predictions. *Journal of Personality and Social Psychology, 70,* 1231-1240.

Haas, A. (1979). Male and female spoken language differences: Stereotypes and evidence. *Psychological Bulletin, 86*(3), 616-626. doi:10.1037/0033-2909.86.3.616

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature, 450*(7169), 557-559.

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: an anger superiority effect. *Journal of Personality and Social Psychology, 54*(6), 917-924.

Heaton, H., & Nygaard, L. (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.

Heblich, S., Lameli, A., & Riener, G. (2015). The effect of perceived regional accents on individual economic behavior: A lab experiment on linguistic performance, cognitive ratings and economic decisions. *PLoS one, 10*(2), e0113475.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*(5), 3099-3111.

Hillenbrand, J. M., & Nearey, T. M. (1999). Identification of resynthesized /hVd/ utterances: Effects of formant contour. *The Journal of the Acoustical Society of America*, *105*(6), 3509-3523.

Hiraga, C. Y., Garry, M. I., Carson, R. G., & Summers, J. J. (2009). Dual-task interference: attentional and neurophysiological influences. *Behavioural Brain Research*, *205*(1), 10-18.

Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology*, *33*(3), 307-316.

Honorof, D. N., Weihing, J., & Fowler, C. A. (2011). Articulatory events are imitated under rapid shadowing. *Journal of Phonetics*, *39*(1), 18-38.

Jaeger, T.F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434-446.

Janik, V. M. (2000). Whistle matching in wild bottlenose dolphins (Tursiops truncatus). *Science*, *289*(5483), 1355-1357.

Jilka, M. (2009). Assessment of phonetic ability. In G. Dogil and M. S. Reiterer (Eds.), *Language talent and brain activity: Volume 1 of trends in applied linguistics* (pp. 257-278). Berlin: de Gruyter.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). The Big Five Inventory – Versions 4a and 54. Berkeley, CA: Institute of Personality and Social Research, University of California, Berkeley.

Kappes, J., Baumgaertner, A., Peschke, C., & Ziegler, W. (2009). Unintended imitation in nonword repetition. *Brain and Language*, *111*(3), 140-151.

Kendall, T. & Thomas, E. R. (2015). Vowels: Vowel manipulation, normalization, and plotting. R package version 1.2-1. (https://cran.r-project.org/package=vowels).

Keppel, G. & Wickens, T. D. (2004). *Design & Analysis: A researcher's handbook*. Upper Saddle River, NJ: Pearson Prentice Hall.

Kim, M. (2011). Phonetic convergence after perceptual exposure to native and nonnative speech: Preliminary findings based on fine-grained acoustic-phonetic measurement. In *Proceedings of the 17th International Congress of Phonetic Sciences*, 1074-1077.

Kim, M. (2012). Phonetic accommodation after auditory exposure to native and nonnative speech. PhD dissertation. Northwestern University.

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, *2*(1), 125-156.

Kinzler, K. D., Shutts, K., DeJesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition*, *27*(4), 623-634.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*(4), 352-358.

Kurzius, E. (2015). The extraverted chameleon: Personality's effects on mimicry of verbal behavior. *Journal of Individual Differences*, *36*(2), 80-86.

Kučera H., & Francis W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Kugler, K. C., Trail, J. B., Dziak, J. J., & Collins, L. M. (2012). *Effect coding versus dummy coding in analysis of data from factorial experiments*. University Park, PA: The Methodology Center, Pennsylvania State University.

Kühn, S., Müller, B. C., van Baaren, R. B., Wietzker, A., Dijksterhuis, A., & Brass, M. (2010). Why do I like you when you behave like me? Neural mechanisms mediating positive consequences of observing someone being imitated. *Social Neuroscience*, *5*(4), 384-392.

Labov, W., Ash, S., & Boberg, C. (2006). *Atlas of North American English: Phonology and phonetics*. Berlin: Mouton.

Latu, I. M., Stewart, T. L., Myers, A. C., Lisco, C. G., Estes, S. B., & Donahue, D. K. (2011). What we "say" and what we "think" about female managers: Explicit versus implicit associations of women with success. *Psychology of Women Quarterly*, *35*(2), 252-266.

Laver, J., & Trudgill, P. (1979). Phonetic and linguistic markers in speech. In K. R. Scherer & H. Giles (Eds.), *Social Markers in Speech* (pp. 1-32). Cambridge, UK: Cambridge University Press.

Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: reliability and validity of the Empathy Quotient. *Psychological Medicine*, *34*(5), 911-920.

Lemasson, A., & Hausberger, M. (2004). Patterns of vocal sharing and social dynamics in a captive group of Campbell's monkeys Cercopithecus campbelli campbelli). *Journal of Comparative Psychology*, *118*(3), 347.

Lemasson, A., Jubin, R., Masataka, N., & Arlet, M. (2016). Copying hierarchical leaders' voices? Acoustic plasticity in female Japanese macaques. *Scientific Reports*, *6*(21289), 1-7.

Lewandowski, E. M., & Nygaard, L. C. *under review*. Vocal alignment to native- and non-native speakers of English.

Lewandowski, N. (2009). Sociolinguistic factors in language proficiency: phonetic convergence as a signature of pronunciation talent. In G. Dogil and M. S. Reiterer (Eds.), *Language talent and brain activity: Volume 1 of trends in applied linguistics* (pp. 257-278). Berlin: de Gruyter.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1-36.

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science*, *36*(8), 1404-1426.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1-36.

Lupfer, M. B., Clark, L. F., & Hutchinson, H. W. (1990). Impact of context on spontaneous trait and situational attributions. *Journal of Personality and Social Psychology*, *58*, 239-249.

Manson, J. H., Bryant, G. A., Gervais, M. M, & Kline, M. A. (2013). Convergence of speech rate in conversation predicts cooperation, *Evolution and Human Behavior, 34*, 419-426.

Marquez, C. S., Ceux, T., & Wenderoth, N. (2011). Attentional demands of movement observation as tested by a dual task approach. *PLoS One*, *6*(11), e27292.

Marshall, A. J., Wrangham, R. W., & Arcadi, A. C. (1999). Does learning affect the structure of vocalizations in chimpanzees?. *Animal Behaviour*, *58*(4), 825-830.

Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 676-684.

Mattys, S. L., Barden, K., & Samuel, A. G. (2014). Extrinsic cognitive load impairs low-level speech perception. *Psychonomic Bulletin & Review*, *21*(3), 748-754.

Mattys, S. L., & Palmer, S. D. (2015). Divided attention disrupts perceptual encoding during speech recognition. *Journal of the Acoustical Society of America*, 137(3), 1464-1472.

Mattys, S. L., & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, *65*(2), 145-160.

McGuire, A. B., Gillath, O., & Vitevitch, M. S. (2016). Effects of mental resource availability on looming task performance. *Attention, Perception, & Psychophysics*, *78*(1), 107-113.

McKirnan, D. J., & Hamayan, E. V. (1984). Speech norms and attitudes toward outgroup members: A test of a model in a bicultural context. *Journal of Language and Social Psychology*, *3*(1), 21-38.

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality*, *40*(4), 525-543.

Mehrabian, A., Young, A. L., & Sato, S. (1988). Emotional empathy and associated individual differences. *Current Psychology*, *7*(3), 221-240.

Meltzoff, A. N., & Decety, J. (2003). What imitation tells us about social cognition: A rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *358*(1431), 491-500.

Meyer, D., Zeileis, A., Hornik, K., Gerber, F., & Friendly, M. (2016). vcd: visualizing categorical data. R package version 1.4-3. (https://CRAN.R-project.org/package=vcd).

Miller, R. M., Sanchez, K., & Rosenblum, L. D. (2013). Is speech alignment to talkers or tasks?. *Attention, Perception, & Psychophysics*, *75*(8), 1817-1826.

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 630-633.

Mitani, J. C., Hunley, K. L., & Murdoch, M. E. (1999). Geographic variation in the calls of wild chimpanzees: A reassessment. *American Journal of Primatology 47*(2), 133-151.

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, *109*(1), 168-173.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132*(2), 297-326.

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*(3), 289-306.

Na, J. & Kitayama, S. (2011). Trait inference is culture-specific: Behavioral and neural evidence. *Psychological Science, 22*(8), 1025-1032.

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology, 21*(4), 422-432.

Neuman, R., & Strack, F. (1998). "Mood contagion": The automatic transfer of mood between persons, *Journal of Personality and Social Psychology, 79*(2), 211-223.

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes, 45*(3), 211-236.

Nielsen, K. (2014). Phonetic imitation by young children and its developmental changes. *Journal of Speech, Language, and Hearing Research*, *57*(6), 2065-2075.

Nygaard, L. C., & Lunders, E. R. (2002). Resolution of lexical ambiguity by emotional tone of voice. *Memory & Cognition*, *30*(4), 583-593.

Nygaard, L. C., Burt, S. A., & Queen, J. S. (2000). Surface form typicality and asymmetric transfer in episodic memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1228-1244.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*(1), 42-46.

Nygaard, L. C., & Queen, J. S. (2008). Communicating emotion: Linking affective prosody and word meaning. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(4), 1017-1030.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*(11), 776-783.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, *25*(1), 46-59.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 309-328.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382-2393.

Pardo, J. S. (2013). Measuring phonetic convergence in speech production. *Frontiers in Psychology*, *4*(559), 183-195.

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics*, *40*(1), 190-197.

Pardo, J. S., Jay, I. C., Hoshino, R., Hasbun, S. M., Sowemimo-Coker, C., & Krauss, R. M. (2013a). Influence of role-switching on phonetic convergence in conversation, *Discourse Processes*, *50*(4), 276-300.

Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception, & Psychophysics*, *72*(8), 2254-2264.

Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., & Lewandowski, E. (2013b). Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Experimental Psychology: Human Perception & Performance, 69(*3), 183-195.

Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics*, *79*(2), 637-659.

Park-Taylor, J., Ng, V., Ventura, A. B., Kang, A. E., Morris, C. R., Gilbert, T., Devika, S., & Androsiglio, R. A. (2008). What it means to be and feel like a “true” American: Perceptions and experiences of second-generation Americans. *Cultural Diversity and Ethnic Minority Psychology*, *14*, 128-122.

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, *54*(1), 547-577.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language, 39*, 633-651.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169-190.

Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, *4*(2-3), 203-228.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), 329-347.

Postma-Nilsenová, M., Mol, L., & Kamoen, N. (2013). If you repeat your interlocutor's syntactic structure, you are likely to repeat her pronunciation, too. In R. In R. Fernandes & A. Isard (Eds.), *Proceedings of SemDial, The 17th Workshop on the Semantics and Pragmatics of Dialogue* (pp. 112-120), Amsterdam: University of Amsterdam.

Prinz, W. (1990). A common coding approach to perception and action. In *Relationships between perception and action,* (pp. 167-201). Berlin: Springer.

Psychology Software Tools, Inc. [E-Prime 2.0]. (2002). Retrieved from http://www.pstnet.com.

Ramos, T., Garcia-Marques, L., Hamilton, D. L., Ferreira, M., & Van Acker, K. (2012). What I infer depends on who you are: The influence of stereotypes on trait and situational spontaneous inferences. *Journal of Experimental Social Psychology*, *48*(6), 1247-1256.

Ramos, T., Oliveira, M., Santos, A. S., Garcia-Marques, L., & Careiro, P. (2016). Evaluating young and old faces on social dimensions: Trustworthiness and dominance. *Psicológia, 37*, 169-185.

Ramos-Fernández, G. (2005). Vocal communication in a fission-fusion society: Do spider monkeys stay in touch with close associates?. *International Journal of Primatology*, *26*(5), 1077-1092.

Raveh, D., & Lavie, N. (2015). Load-induced inattentional deafness. *Attention, Perception, & Psychophysics*, *77*(2), 483-492.

Ruch, H., Zürcher, Y., & Burkart, J. M. (2017). The function and mechanism of vocal accommodation in humans and other primates. *Biological Reviews*. Doi: 10.1111/brv.12382.

Rukstalis, M., Fite, J. E., & French, J. A. (2003). Social change affects vocal structure in a callitrichid primate (Callithrix kuhlii). *Ethology*, *109*(4), 327-340.

Ryan, E. B. (1983). Social psychological mechanisms underlying native speaker evaluations of non-native speech. *Studies in Second Language Acquisition, 5,* 148-159.

Ryan, E. B., Carranza, M. A., & Moffie, R. W. (1977). Reactions toward varying degrees of accentedness in the speech of Spanish-English bilinguals. *Language and Speech*, *20*(3), 267-273.

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics, 25*, 421-436.

Schulz, C., Mothes-Lasch, M., & Straube, T. (2013). Automatic neural processing of disorder-related stimuli in social anxiety disorder: Faces and more. *Frontiers in Psychology*, *4*(282), 1-16.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*(1), 1-66.

Seitz, A. R., Protopapas, A., Tsushima, Y., Vlahou, E. L., Gori, S., Grossberg, S., & Watanabe, T. (2010). Unattended exposure to components of speech sounds yields same benefits as explicit auditory training. *Cognition*, *115*(3), 435-443.

Seyfarth, R. M., & Cheney, D. L. (2013). Affiliation, empathy, and the origins of theory of mind. *Proceedings of the National Academy of Sciences*, *110*(2), 10349-10356.

Shockley, K., Baker, A. A., Richardson, M. J., & Fowler, C. A. (2007). Articulatory constraints on interpersonal postural coordination. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 201-208.

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, *66*(3), 422-429.

Sidaras, S. (2011). Hearing what you expect to hear: The interaction of social and cognitive mechanisms underling vocal accommodation. PhD dissertation. Emory University.

Silk, J. B. (2014). Evolutionary Perspectives on the Links Between Close Social Bonds, Health, and Fitness. In M. Weinstein M and M. A. Lane (Eds.), *Comparative Biodemography: A Collection of Papers*. Washington, DC: National Academies Press.

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2016). Afex: analysis of factorial experiments. (https://CRAN.R-project.org/package=afex).

Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain and Language*, *164*, 77-105.

Smith, D. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgments of speaker size, sex, and age. *The Journal of the Acoustical Society of America*, *118*(5), 3177-3186.

Smolker, R., & Pepper, J. W. (1999). Whistle convergence among allied male bottlenose dolphins (Delphinidae, Tursiops sp.). *Ethology*, *105*(7), 595-617.

Snowdon, C. T., & Elowson, A. M. (1999). Pygmy marmosets modify call structure when paired. *Ethology*, *105*(10), 893-908.

Snyder, M., Tanke, E. D., & Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology, 35*, 656-666.

Solanki, V., Vinciarelli, A., Stuart-Smith, J., & Smith, R. (2016). When the game gets difficult, then it is time for mimicry. In A. Esposito et al. (Eds.) *Recent Advances in Nonlinear Speech Processing* (pp. 247-254). New York, NY: Springer-Verlag.

Soliz, J., & Giles, H. (2014). Relational and identity processes in communication: A contextual and meta-analytical review of Communication Accommodation Theory. *Annals of the International Communication Association*, *38*(1), 107-144.

Srull, T. K., & Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology, 37*(10), 1660-1672.

Stiff, J. B., Dillard, J. P., Somera, L., Kim, H., & Sleight, C. (1988). Empathy, communication, and prosocial behavior. *Communications Monographs*, *55*(2), 198-213.

Street, R. L. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research*, *11*(2), 139-169.

Subiaul, F., & Schilder, B. (2014) Working memory constraints on imitation and emulation. *Journal of Experimental Child Psychology*, 128, 190-200.

Subiaul, F., Winters, K., Krumpak, K. & Core, C. (2016). Vocal overimitation in preschool-age children. *Journal of Experimental Child Psychology, 141*, 145-160.

Sugiura, H. (1998). Matching of acoustic features during the vocal exchange of coo calls by Japanese macaques. *Animal Behaviour*, *55*(3), 673-687.

Titze, I. R., Baken, R.J. Bozeman, K. W., Granqvist, S., Henrich, N., Herbst, C. T., Howard, D. M., Hunter, E. J., Kaelin, D., Kent, R. D., Löfqvist, A., McCoy, S., Miller, D. G., Noé, H., Scherer, R. C., Smith, J. R., Story, B. H., Švec, J. G., Ternström, S., & Wolfe, J. (2015). Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization. *Journal of the Acoustical Society of America*, *137*, 3005-3007.

Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, *83*(5), 1051-1065.

Tombu, M., & Jolicœur, P. (2003). A central capacity sharing model of dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(1), 3-18.

Trudgill, P. (2008). Colonial dialect contact in the history of European languages: On the irrelevance of identity to new-dialect formation. *Language in Society*, *37*(2), 241-254.

Turk, D. J., Brady-van den Bos, M., Collard, P., Gillespie-Smith, K., Conway, M. A., & Cunningham, S. J. (2013). Divided attention selectively impairs memory for self-relevant information. *Memory & Cognition*, *41*(4), 503-510.

Tuten, D. N. (2008). Identity formation and accommodation: Sequential and simultaneous relations. *Language in Society*, *37*(2), 259-262.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207-232.

Uleman, J. S., Newman, L., & Winter, L. (1992). Can personality traits be inferred automatically? Spontaneous inferences require cognitive capacity at encoding. *Consciousness and Cognition*, *1*, 77-90.

Unsworth N., Heitz R. P., Schrock J. C., Engle R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498-505.

Van Baaren, R. B., Holland, R. W., Kawakami, K., & Van Knippenberg, A. (2004). Mimicry and prosocial behavior. *Psychological Science*, *15*(1), 71-74.

Van Baaren, R., Janssen, L., Chartrand, T. L., & Dijksterhuis, A. (2009). Where is the love? The social aspects of mimicry. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1528), 2381-2389.

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.

Van Leeuwen, M. L., van Baaren, R. B., Martin, D., Dijksterhuis, A., & Bekkering, H. (2009). Executive functioning and imitation: Increasing working memory load facilitates behavioural imitation. *Neuropsychologia*, *47*(14), 3265-3270.

Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., & Dolan, R. J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nature Neuroscience, 7*(11), 1271-1278.

Walker, A., & Campbell-Kibler, K. (2015). Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task. *Frontiers in Psychology, 6,* 1-18.

Watson, S. K., Townsend, S. W., Schel, A. M., Wilke, C., Wallace, E. K., Cheng, L., West, V., & Slocombe, K. E. (2015a). Vocal learning in the functionally referential food grunts of chimpanzees. *Current Biology*, *25*(4), 495-499.

Watson, S. K., Townsend, S. W., Schel, A. M., Wilke, C., Wallace, E. K., Cheng, L., West, V., & Slocombe, K. E. (2015b). Reply to Fischer et al. *Current Biology*, *25*(4), R1030-R1031.

Wells, B. M., Skowronski, J. J., Crawford, M. T., Scherer, C. R., & Carlston, D. E. (2011). Inference making and linking both require thinking: Spontaneous trait inference and spontaneous trait transference both rely on working memory capacity. *Journal of Experimental Social Psychology, 47*(6), 1116-1126.

Wickens, C. (1981). Processing resources in attention. *Processing Resources in Attention and Workload: Dual Task Performance and Workload Assessment*, (pp. 3-34). Bristol, PA: Taylor & Francis.

Wigboldus, D. H., Dijksterhuis, A., & van Knippenberg, A. (2003). When stereotypes get in the way: Stereotypes obstruct stereotype-inconsistent trait inferences. *Journal of Personality and Social Psychology*, *84*(3), 470-484.

Wigboldus, D. H., Sherman, J. W., Franzese, H. L., & Knippenberg, A. V. (2004). Capacity and comprehension: Spontaneous stereotyping under cognitive load. *Social Cognition, 22*(3), 292-309.

Williams, P. (1998). *Representational Organization of Multiple Exemplars of Object Categories*. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.5.8336.pdf.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*(2), 237-252.

Winter, L., Uleman, J. S., & Cunniff, C. (1985). How automatic are social judgments? *Journal of Personality and Social Psychology, 49*, 904-917.

Wolak, M. (2016). ICC: Facilitating estimation of the Intraclass Correlation Coefficient. R package version 2.3.0. (http://github.com/matthewwolak/ICC).

Yu, A. C. L., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and “autistic” traits. *PloS one*, *8*(9), e74746.

Zürcher, Y., & Burkart, J. M. (2017). Evidence for dialects in three captive populations of common marmosets (Callithrix jacchus). *International Journal of Primatology*, *38*(4), 780-793.

Appendix A: Full Word Lists

Experiment 1 Word List			
Easy Words		Hard Words	
balm	live	bean	main
cause	lose	chat	mall
curve	mouth	cheer	mat
death	move	chore	mitt
deep	path	cod	moat
dog	peace	dame	mum
down	pool	den	pat
fig	pull	doom	pet
firm	put	dune	rat
fool	real	goat	rhyme
gave	rough	hack	rim
girl	shall	hash	rum
hung	size	hum	sane
join	soil	kin	toot
judge	theme	kit	wail
king	thick	knob	white
league	thought	lace	whore
learn	work	lice	wrong

Experiment 2 Word List			
High Frequency		Low Frequency	
anon	monster	abscond	meager
apart	motif	arson	menace
assume	movement	banal	misdeal
cannot	movie	bardic	miso
center	neither	bazaar	moonlit
connect	never	beady	narwhal
dealer	people	benzene	nocturne
demo	percent	burble	octet
disease	placement	cadet	offend
duty	reduce	cahoots	peaky
event	regard	chenille	poodle
female	remove	demean	preheat
follow	resume	denude	raccoon
forget	review	divest	seduce
garbage	robot	draftee	senile
indeed	special	ensconce	sensual
instead	student	foment	snooty
into	super	gavotte	stocking
involve	topic	ghoulish	supine
issue	unique	harlot	vamoose
machine	vendor	imam	vegan
marvel	volume	immense	vengeance
meeting	workshop	kazoo	volley
member	zero	meadow	voodoo

Appendix B: Character Photo and Narrative Ratings

		
Attractiveness	3.94	3.89
Dominance	4.98	4.67
Gender Typicality	6.38	6.07
Likeability	4.98	4.88
Perceived Age	23.96	23.84
Trustworthiness	5.27	5.70

Ratings collected by Ramos, Oliveira, Santos, Garcia-Marques, & Carneiro (2016), originally published in Psicológia. Reproduced under the terms of a Creative Commons Attribution-NonComm-NoDerivs 4.0 License.

Bad: Ed is disrespectful to everyone, but especially to members of the opposite sex. He is also quite racist and likes to brag about having committed violence against minorities. It's simply because they are different from him.

Good: After I lost my rent money, I called my landlord during my break at work. Apparently, Joe overheard me pleading for time. The next day, he brought me the money I needed. He didn't have much himself.

	Bad Narrative	**Good Narrative**
Attractiveness	1.95	3.6
Character*	6.15	5.65
Cheating	3.55	3
Competence†	3.3	3.2
Confidence	6	5.8
Intensity	5.45	3.95
Trustworthiness	4.8	4.75
Valence	1.65	6.15
Warmth	5.3	5.3
Weirdness	3.3	3.3

* The "Character" score indicates how revealing the narrative is about a person's character, regardless of whether is is good character or bad character. "Valence" is an indication of whether the character is good or bad.

† "Competence" indicates how revealing the narrative is about a person's level of competence, regardless of whether it is high or low.

Character narratives written and normed by Beal (2018), reproduced with permission.