Spring 2017

Cooperation via Communication: Influencing Vocal Alignment in Conversation

Elliot A. Pollack
Trinity College, Hartford Connecticut, elliot.pollack@trincoll.edu

Cooperation via Communication:

Influencing Vocal Alignment in Conversation

Elliot Pollack

Faculty Advisor: Elizabeth Casserly, Ph. D

Department of Psychology

Trinity College

2016-2017

Abstract

Alignment of human behavior is a well-documented phenomenon, however, the factors which influence its direction and magnitude are not firmly established. Conversational partners align on a variety of speech factors including word choice, syntax, and rate of speech. The present study examines factors which lead to alignment of fundamental frequency (F0), colloquially known as pitch. Subjects (Speakers) complete a puzzle task which requires them to communicate with a partner (Model). The Model's F0 is manipulated to either converge towards or diverge from that of the Speaker, whereas a control condition does not change the Model voice. The Speaker is recorded throughout the interaction (Task); baseline (Pre-task) and final (Post-task) recordings are also taken. Speakers' F0 is measured at each time-period to determine the direction and magnitude of alignment. In a separate session, naïve subjects (Listeners) assess the similarity between the Speaker's speech over time and the Model. A personality survey examines which factors serve as reliable predictors of alignment. Speakers are found to deviate from the Model in F0 during the interaction, however, are perceived by Listeners to mimic the Model over time in a holistic measure. These findings are consistent regardless of the Model's direction of alignment. Speakers are rated as becoming more like the Model when this partner diverges as opposed to converges. The personality factor survey shows that Openness predicts alignment. Specifically, greater Openness predicts less perceived similarity. None of the other personality factors (conscientiousness, extraversion, agreeableness, neuroticism) are found to share a significant relationship with alignment behavior. Alignment between any two time-periods throughout the experiment predicts alignment with the third. The discrepancy between Speakers' divergence in the acoustic measure and their rated convergence in the perceptual measure reveals a potential hierarchy of speech factors that we use to assess alignment.

Introduction

Overview

Alignment is salient in human behavior. Human beings are sociable animals and we are interdependent on one another for food, shelter, and companionship. We engage in the use of simple and complex behaviors alike in response to the actions of others with whom we interact. Our actions—or reactions to the behavior we observe around us—are largely an effortless process. In many cases, such behaviors are automatic and outside of conscious awareness.

The scientific literature contains numerous examples of alignment in behavior. Richardson and colleagues (2007) examined alignment between randomly assigned pairs seated in adjacent rocking chairs. The researchers found that participants would synchronize the rate at which they rocked so that the pair would move forward and backward at the same time. This alignment of rocking behavior was observed even between members of pairs who were instructed to rock at their individually preferred tempo, as opposed to instructions explicitly requiring the pair to rock in synchronization.

Our tendency to align is automatic and can be extended to incidental behaviors that occur without us being fully aware of our engagement. Chartrand and Bargh (1999) researched the “Chameleon effect,” in which people mimic the behavior of those around them. They discovered that participants would mimic the bodily actions, such as face rubbing and foot shaking, as well as the facial expressions of confederates, even though these behaviors were in no way related to the picture-describing task at hand. The researchers took these results as support for a link between prior perception and subsequent behavior. The take-home message is that alignment is both natural and normal. This phenomenon often occurs in behaviors of which we are unaware that can be observed in everyday life. Observe at a lecture whether people in attendance align in behaviors such as note taking, head scratching, and clapping at the end of the presentation. The present study examines alignment in speech and places a magnified focus on pitch.

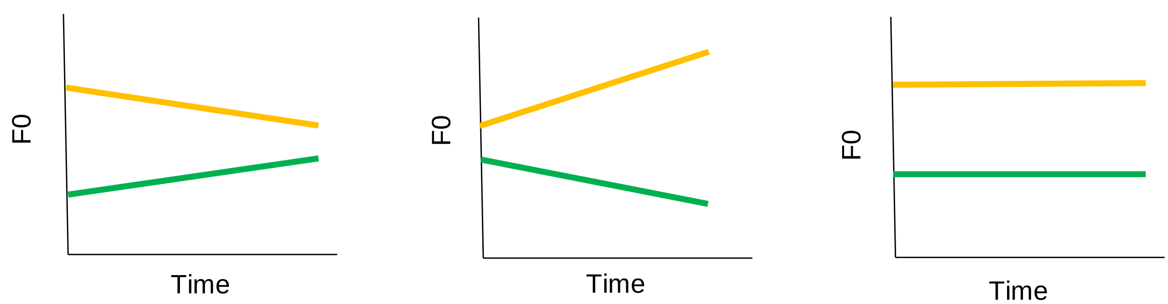
Types of Alignment

The term *alignment*, in this study, refers to a change in behavior on the part of one person in response to the observed behavior of a partner in an interacting dyad. This definition is unique to the current paper, and is more inclusive than the term *accommodation* which is used extensively in the literature to refer to imitative behavior (e.g. Babel & Bulatov, 2011; Gregory & Webster, 1996; Gijssels et al., 2016). For this study, *alignment* itself does not specify how the behavior changes. The current investigation examines three types of alignment which each define a unique direction of the change in behavior.

Convergence (see Figure 1): One person changes their behavior in such a way that it becomes closer to that of their partner than before the interaction along some dimension. The aforementioned studies depict this type of alignment. Matching the rate of rocking in adjacent chairs (Richardson et al., 2007) and mimicking the body language and facial expressions of a partner (Chartrand & Bargh, 1999) both consist of people increasing the similarity of their behavior in relation to their interactive partner. In speech, this type of alignment is evident in the research of Pardo and colleagues (2012), which found that the perceived similarity in the speech of male undergraduate roommates increased over the course of the semester.

Divergence (see Figure 2): The opposite of convergence, namely, a change in behavior that results in an action that is further from that of a partner than before the interaction along some dimension. A speech related example can be found in a study conducted by Chen and colleagues (2010), in which participants received manipulated auditory feedback of their own speech. The researchers found that speakers were more likely to produce opposing (diverging) responses—changes in their pitch in the opposite direction of the shift—than following (converging) responses. In this case, the sole participant's sensory feedback filled the role of the interactive partner that is used in more conventional studies of vocal alignment.

Neutral (see Figure 3): A third type is not alignment per se, but an alternative to convergence and divergence. There is no net change in the difference between individuals along some dimension of behavior. Fusaroli and colleagues (2012) illustrated two types of alignment for speech which may help to clarify this type. The first, local linguistic alignment, can be defined as one person adapting to the way that another person talks. This type has no local linguistic alignment. The second, global linguistic convergence, is the alignment of a set of shared expressions between conversational partners. This global feature is also absent in this type of alignment because the individual is not making a change in either direction in response to their partner's behavior.



Figures 1 (left), *2* (center), *3* (right). Two talkers exhibit converging (left), diverging (middle), and neutral (right) behavior in F0 (pitch) alignment across time.

Speech

There are several ways in which alignment manifests in human speech behavior. The process can be seen in two different levels of speech: discrete and continuous (Gijssels et al., 2016). These two categories identify how the feature of speech is measured. Each contains several features of vocal behavior—such as word choice, syntax, and rate of speech—which are subject to alignment. Many of these features have been examined in prior research.

Discrete. Features of speech which fall under this category are measured with a finite set of values. Examples include word choice (e.g. choose between two synonyms such as *tasty* or *delicious*) and syntax (e.g. use either active or passive voice). Alignment in a discrete dimension measures whether speakers use language which either mimics their partner or deviates in binary fashion. For word choice, participants converge when they utilize the same words that their partner just produced—local linguistic alignment—and a pair of participants may converge towards one another by adopting a shared set of expressions, global linguistic convergence (Fusaroli et al., 2012). Convergence of syntax requires that a member of a conversational dyad utilize the same grammatical structures as their partner (e.g. Kaschak et al., 2011). Divergence in these features of speech would entail progression of the interaction with increasing differences in word choice and syntax. In both cases, the direction of alignment depends upon whether speakers use the same or different language, with no middle ground.

Continuous. This category of speech features elements which are measured quantitatively with an infinite set of possible values. Examples include rate of speech (faster or slower) and pitch (as a frequency, measured in Hertz). Alignment in this dimension measures whether the quantitative difference between speakers changes over the course of an interaction. Convergence in rate of speech entails interlocutors moving closer to matching one another in the number of words/syllables produced per unit of time, whereas divergence entails a gap between the individuals' production speed as the interaction progresses. Alignment in pitch is assessed in a similar fashion. Convergence describes the process of a speaker moving closer towards the pitch of a partner as the pair interacts; divergence consists of a speaker changing their pitch so that it is further from matching that of their partner. A neutral response would consist of neither an increase nor decrease in the pitch difference between interlocutors. The degree of alignment in each case can be measured in both direction and magnitude.

Common Priming Mechanism

A question arises in how this process of alignment occurs. Pickering and Garrod (2004) proposed the influential Interactive Alignment Model which attempts to answer this question. The model suggests that alignment for the various features of speech occurs through a priming mechanism. Interlocutor A's words, syntax, and rate of speech—among other features—are perceived by their partner, interlocutor B. The exact language chosen by interlocutor A is subsequently activated in the appropriate regions of interlocutor B's brain. Now the language which A uses is primed in B's cognitive network so that when B speaks, A's language is closer to the activation threshold than alternative forms, all else being equal. Therefore, there is an increased likelihood that B will produce the same words, syntax, and rate of speech which were uttered by A. This is a largely automatic process which helps to make the language production process more efficient. Note that Pickering and Garrod (2004) asserted that the priming mechanism is common to both discrete and continuous levels of speech.

The alignment of discrete features of speech is compatible with the Interactive Alignment Model (Kaschat et al., 2011; Fusaroli et al., 2012). Kaschak and colleagues (2011), for example, have shown that syntactic structures such as double object (e.g. *give the dog a treat*) versus prepositional object (e.g. *give a treat to the dog*) were subject to priming in which producing speech in a setting where one form of syntax was used increased the likelihood of using the same grammatical structure when situated in the same environment. In this case, syntactic structures were being primed within a lone speaker. The Interactive Alignment Model suggests that the speech from other speakers should be subject to this exact priming mechanism (Pickering & Garrod, 2004). This appears to be the case; interlocutors have been shown to reuse expressions uttered by their partners and converge on a common vocabulary for task relevant utterances (Fusaroli et al., 2012).

Furthermore, the literature provides support for the alignment of continuous features of speech via the same priming mechanism (Himberg et al., 2015; Bilbous & Krauss, 1988; Namy et al., 2002). Himberg and colleagues (2015), for example, found that interlocutors synchronized their inter-turn intervals (i.e. the time between Speaker A beginning one word and the next, with Speaker B intervening) as they progressed through a story completion task. When considering sex differences in alignment, male dyads converged on utterance length and pauses, whereas female dyads also converged on the total number of words uttered and interruptions (Bilbous & Krauss, 1988). Namy and colleagues (2002) asserted that the sex discrepancy in vocal accommodation is due to differences in the perceived characteristics of speech that are evident to men and women. This ties back to the Interactive Alignment Model, as speakers were converging on features of speech that they were most readily able to perceive, which is necessary for priming and subsequent alignment (Pickering & Garrod, 2004).

While the Interactive Alignment Model has significant clout in the psycholinguistic community, the theory does not receive unanimous support. Gijssels and colleagues (2016) challenge the generalizability of Pickering and Garrod’s (2004) model to continuous features of speech. Specifically, they do not believe that the priming mechanism adequately accounts for the alignment of continuous features of speech such as pitch. The researchers argued that if pitch were primed, then there should be an increase in alignment as the conversation progressed and effects that last beyond the end of the interaction. But when the researchers examined pitch alignment trajectories and their maintenance after the experimental interaction, they found exactly the opposite (Gijssels et al., 2016). Namely, alignment in pitch failed to increase in strength over the course of the interaction and terminated instantly once the conversation ceased. Both criteria for a priming mechanism failed to manifest in pitch alignment.

The challenge to Pickering and Garrod’s (2004) model is significant, however, it should be taken with a grain of salt. The Interactive Alignment Model still has much support in explaining the process of alignment in continuous dimensions of speech, such as pitch (e.g. Babel & Bulatov, 2011; Behroozmand et al., 2012; Jones & Munhall, 2000). But it should be noted that further research is needed on this front and that future studies should work to address potential alternative models. The Interactive Alignment Model provides one potential mechanism by which the alignment process occurs, but it is by no means the only influence acting upon this phenomenon. For example, there is a large body of literature suggesting that social motivation plays a significant role in alignment.

Purpose of Alignment

The Interactive Alignment Model offers a feasible proposal to explain *how* the alignment process occurs, but does little to explain *why* we align our vocal behavior. To answer this question, the psycholinguistic community largely turns to the Communication Accommodation Theory proposed by Giles and colleagues (1991). The theory considers alignment to be an evolutionary response designed to effectively facilitate social interactions by changing social distance (i.e. become either closer or further from being associated with another person). In turn, this promotes more effective communication of social goals and stances between the speaker and listener. While Pickering and Garrod’s (2004) model primarily addressed convergence, Giles and colleagues (1991) considered both convergence and divergence with their theory. Convergence emerges as an effort to decrease social distance. Attraction and group cohesion promote converging behavior. On the contrary, divergence serves to increase social distance. This is evident when distinction is prioritized. It is important to note that the theory allows for convergence to occur for some features of speech while other dimensions may experience divergence during the same interaction. The literature is replete with investigations of the role that social factors play in bidirectional alignment.

One social factor which appears to play a major role in alignment is the speaker’s attitude towards the listener. A speaker’s automatic and social biases concerning their listener or conversational partner predict the degree of alignment, for example (Babel, 2010). Specifically, positive biases are associated with greater convergence (Babel, 2010). Similarly, speech convergence in college roommates was moderated by feelings of closeness between the dyad members (Pardo et al., 2012). The social factors even extend beyond human-to-human interactions. People have been shown to align with computers (Lee, 2010). Computers perceived to sound more human-like instilled more converging behavior from participants (Lee, 2010). The takeaway from these studies is that alignment, convergence in these examples, was fueled by the desire to decrease social distance with those who the speaker deemed more attractive (Giles et al., 1991). Furthermore, once alignment is initiated, it can begin a cycle. Perceived alignment on the part of the listener makes the speaker appear more favorable which, in turn, leads to the listener reciprocating the effort made by the initial speaker (Giles et al., 1973).

Another area which is implicated in alignment is the realm of social status, societal roles, and hierarchy. A cleverly designed study conducted by Gregory and Webster (1996) examined accommodation that took place between a television news host, Larry King, and his guests. Lower status guests—as rated by naive third party participants—were found to converge in pitch to a greater degree than higher status guests. Additionally, lower status guests would accommodate to King, whereas King would converge towards higher status individuals. This result may be attributed to the lower status interlocutor needing to gain the approval of the higher status partner (Giles et al., 1991). A similar hierarchical relationship can be observed in the therapeutic environment, albeit to a lesser degree. Reich and colleagues (2014) found that therapists were more likely than not to lead the pitch shifts. While this trend was nonsignificant, it alludes to a potential hierarchical relationship with the therapist in the dominant role and the patient being relatively submissive.

Societal roles for the two sexes are another factor to consider for alignment. There exists a discrepancy in the literature as to which sex accommodates to a greater degree; some studies depict male speakers exhibiting greater alignment (e.g. Pardo, 2006) whereas others identify female speakers as the greater aligners (e.g. Namy et al., 2002). Additionally, there are between-group differences concerning which features of speech were subject to convergence and divergence (Bilous & Krauss, 1988). To attempt to explain this discrepancy, Pardo (2006) proposed that the individuals comprising the two sexes may have differently interpreted the task's instructional roles as either dominant or submissive.

Finally, a potentially significant factor, or set of factors, that can serve as a valid predictor of alignment is the Big Five Personality Trait assessment. The five traits—openness, conscientiousness, extroversion, agreeableness, neuroticism—may play a significant role in accommodation (Yu et al., 2013). The current literature lacks sufficient investigation of this relationship, however, the research that has been conducted suggests that openness is a significant predictor of alignment in voice onset time (Yu et al., 2013). The interaction of all five traits with one another and speech alignment remains unknown.

The Present Study

The primary aim of the current investigation is to further our understanding of the causes and dynamics of speech alignment. The alignment of F0—a continuous speech characteristic—was measured to address three specific areas of interest:

1. *How does the alignment of one member of a dyad influence the alignment of the conversational partner?*

   It is difficult to determine causation in real-life social interactions from experimental conversations. For example, in the Gregory and Webster (1996) study, alignment may have occurred for several reasons. Perhaps lower status individuals noticed that King was not converging so they accommodated their host. Alternatively, King may have noticed that his lower status guests were converging so there was no need to accommodate on his part. Perhaps the observed convergence was a mix of both causes.

2. *Does our assessment of alignment change based upon whether we look through an acoustic or perceptual lens?*

   Both acoustic and perceptual analyses are important for assessing alignment (Miller et al., 2013; Pardo et al., 2013). The acoustics provide quantifiable changes to specific dimensions of speech. Complimentarily, the perceptual analysis is more holistic in its assessment of alignment and provides a perspective which is more akin to everyday social interactions.

3. *What personality factors and attitudes towards an interlocutor serve as reliable predictors of alignment?*

   Yu and colleagues (2013) found that greater Openness predicts more convergence in voice onset time. It has yet to be determined whether Openness can predict alignment for other features of speech, or for a holistic assessment. Additionally, Babel (2010) identified positive biases towards a listener or conversational partner as a predictor of greater convergence on the part of the speaker. The current study will assess this relationship with a conversational partner who changes pitch during the interaction.

The experiment designed to address these queries consists of a puzzle which requires cooperation and effective verbal communication between two partners. As the dyad completes the puzzle, the pitch of one speaker's voice is experimentally manipulated in one of three ways with respect to the participant's median F0: convergence, divergence, or neutral (non-shift). All other features of speech are held constant. The direction and magnitude of alignment for the partner who does not make the initial shift is recorded. Prior research verifies the efficacy of puzzle-like tasks for inducing alignment (e.g. Dias & Rosenblum, 2011). Furthermore, cooperative behavior has been shown to correlate with alignment (e.g. Manson et al., 2013). The success of the approach in this investigation yields the ability for one interlocutor to influence the accommodation of their partner by manipulating their own pitch alignment. Assessment of alignment occurs in two phases: pitch (acoustic) and holistic (perception) measurements.

Pitch. This characteristic of speech is what listeners perceive when they encounter the sound that speakers produce through their vocal fold vibration (fundamental frequency, abbreviated as F0). Pitch is a critical component of speech which is unique for each person and changes throughout the day (Cooper & Yanagihara, 1971). This variability allows F0 to be particularly susceptible to alignment behavior. Prior research revealed that the presence of fundamental frequency in an interlocutor's voice led to greater degrees of alignment without manipulating any other features of speech (Babel & Bulatov, 2011). For the initial phase of the current study, speech alignment is operationalized as participants' changes in pitch which occur in response to the shift of their respective partners.

Holistic. Acoustic measurements, while concrete, do not tell the whole story of alignment. Miller and colleagues (2013) as well as Pardo and colleagues (2013) argue that perceptual analysis is required in addition to acoustic analysis because the former is a holistic measure of accommodation. The researchers suggested that this type of assessment is necessary because alignment serves a social function and people in everyday conversation do not parse speech apart into various phonetic dimensions, but rather, they perceive and produce speech in a holistic manner. The downside to this approach is that it is not as "objective" as acoustic analysis in the sense that there is no unit of measurement like there is with pitch (Hz). In the final phase of the current study, naïve raters judge the perceptual similarity of the partners' speech at different stages of the puzzle task to create a more complete picture of alignment.

I hypothesize that the following will result from my experiments:

1. *Alignment of participants depends upon the alignment of their partner. If the partner converges or diverges, then the participant will follow suit. But if the partner is neutral, then the participant will still converge, but to a lesser degree than when the partner converges.*

   When one conversational partner aligns with respect to the other, there may be a change in attitude towards that initial partner and subsequent alignment at the other end of the interaction (Giles et al., 1973). For convergence, this translates into more positive attitudes and a convergent response to match for the other partner. On the other hand, diverging behavior would lead to more negative attitudes and a divergent response on the part of the other interlocutor. As for the neutral condition, with all else equal, Miller and colleagues (2013) found that people converge towards the specific person with whom they are interacting.
2. *The assessment of alignment will remain the same for the acoustic and perceptual measures.* If a common priming mechanism does apply to all features of speech (Pickering & Garrod, 2004), then all features of speech should align in the same direction as each another.
3. *Personality factors and attitudes towards a conversational partner serve as reliable predictors of alignment.* Considering the association between greater Openness and convergence in voice onset time (Yu et al., 2013), greater Openness should be associated with an increase in speech convergence when all features of speech are

taken into account via a holistic assessment. Babel’s (2010) findings hinted that positive perceptions of a partner’s friendliness and cooperativeness should be correlated with converging speech behavior.

Methods

Phase I

Subjects. Phase I was conducted at a New England college with 42 subjects (mean age = 18.8 years, SD = 0.899; female = 24, male = 16, not declared = 2), to be referred to as *Speakers*. All Speakers were undergraduate students at the campus where the study took place. Additionally, Speakers were all self-assessed as fluent in English and reported no hearing impairments at the time of the experiment. A hearing screening confirmed that all Speakers were at 30 dB hearing level or lower at the time of the experiment. Speakers were compensated with either credit for an introductory psychology course or a gift card at the rate of $10 per hour.

Experimental Design. Each Speaker was randomized into one of three conditions (14 in each) as they interacted with their partner to complete the puzzle task.

1. *Convergence*: The Model’s voice was shifted towards the Speaker’s by 10% of the difference between the Model’s median F0 and the Speaker’s median F0, which was calculated from their five Pre-task recordings.
2. *Divergence*: The Model’s voice was shifted away from the Speaker’s by 10% of the difference between the Model’s median F0 and the Speaker’s median F0; again, this was calculated from their five Pre-task recordings.
3. *Neutral*: The Model’s voice was kept constant at his median F0 for all recordings.

Procedure. Speakers were greeted by the experimenter (Model) upon entering the laboratory. Once consent was obtained, Speakers were seated in a sound proof booth (Whisper Room). The experimenter fitted the Speaker with the headphones and head-mounted microphone. In front of the Speaker was a desk with the puzzle board in the front-center section, monitor in the back-center, and puzzle pieces in transparent cups situated on both sides of the puzzle board. The pieces were arranged by color, with each cup containing all possible shapes of that color to be used in the puzzle task. The experimenter informed the Speaker that he will be in a different room and able to communicate with the Speaker through the headset so that they can complete the puzzle task together. This was deceptive; the Model just listened to the Speaker but then transmitted previously recorded speech to the Speaker's headset. The Model exited the booth. The instructional video was played on the monitor in the sound proof booth.

Pre-task. Once the video ended, the Speaker immediately began the Pre-task. Five pairs of colored shapes were presented, with one pair shown at a time. The Speaker described the colored-shape pair as demonstrated in the video. Their speech was recorded using Praat.

Task. The puzzle task followed the Pre-task. Once the Pre-task speech was recorded, the Model used Praat to assess the Speaker's median baseline pitch from the Pre-task.

In all three conditions, the Model began the Task by playing the first recording, which named the first colored-shape to be placed on the board. The Speaker placed the appropriate piece on their puzzle board, and followed by describing the next colored-shape pair. The Model proceeded by playing the next recording, and this continued in alternating fashion until the puzzle was complete. The Speaker's speech was recorded throughout the entire Task using Praat.

Post-task. The Post-task mirrored the Pre-task. The colored shape pairs were the exact same as in the Pre-task. This portion of the experiment served as a final measurement to determine whether alignment persists upon termination of the interaction. The Speaker's speech was recorded using Praat.

Once the Post-task was completed, the Model retrieved the Speaker from the sound-proof booth. The Speaker completed the lifestyle and abbreviated Big-5 Personality Trait survey (see Appendix A). After completing the survey, the Model informed the Speaker about the deception employed in the study, namely that the Model's speech was recorded prior to the experiment in lieu of the live communication which the Speaker was told would take place during completion of the puzzle task.

Materials. The puzzle which Speakers solved for the cooperative task was a poster board which consisted of an 8 x 5 grid with a total of 40 spaces (see Figure 4). Each space contained either a pair of colored-shapes, or a single colored shape followed by a piece of Velcro which must be filled in by the Speaker.

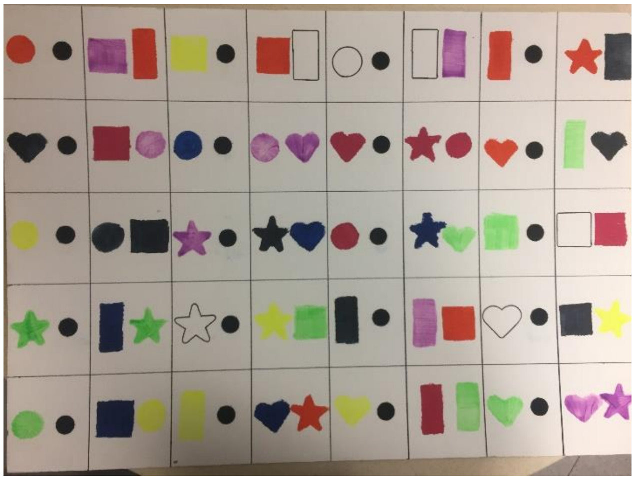


Figure 4. The puzzle board used in Phase I. Speakers would listen and fill in grid spaces with a blank spot (Velcro) and speak aloud the colored-shape pairs in the grid spaces that already had both slots filled.

Puzzle board pieces were wooden and took the form of colored-shapes. Examples include a blue square, green heart, and orange square, among others. Each piece was equipped with a piece of Velcro so that the Speaker may attach it to the puzzle board.

Acoustic stimuli utilized for the puzzle task were speech recordings from the experimenter, to be referred to as the Model. Each utterance was complimentary to the Speakers' puzzle board. The Model recordings revealed to Speakers the correct piece to place on each of the blank Velcro spaces on the puzzle. The form of the Model's utterances was "Next to the *color 1* *shape 1* is a *color 2* *shape 2*". For example, a possible utterance for the Model was "Next to the blue square is a green heart". The acoustic stimuli were presented to Speakers via over-the-ear headphones (Sennhesier HD 280 Professional). The fundamental frequency of the Model's first recording was set to the Model's median F0 (93.4 Hz in this experiment). The remaining 19 Model recordings were manipulated as determined by the experimental condition for that Speaker. There was a total of 20 Model recordings.

The speech of Speakers was recorded during the Pre-task, Task, and Post-task using a head-mounted microphone (AKG C 250 with AKG MPA VL Micro Mic Phantom Adaptor for 9 to 52 Volts). The microphone was worn simultaneously with the headphones.

Visual stimuli for the Pre-task and Post-task included colored-shape pairs which resembled those on the puzzle board. These stimuli were presented on a monitor (Samsung Model S23C350H) with a Microsoft PowerPoint slideshow presentation. It was these stimuli which were the target phrases used for analysis, as will be described below.

An instructional video was shown to the Speakers prior to the Pre-task using the monitor. The video provided a visual demonstration of the puzzle task to be completed. Additionally, the video *de-aligned* the Speaker from the Model (experimenter) prior to the start of the Pre-task. This was accomplished by having someone other than the Model explain the instructions in the video. It was necessary because Speakers may align speech with the Model in the interaction when obtaining consent. Gijssels and colleagues (2016) reported that alignment occurs immediately after the interaction is initiated. The video brought the Speakers' pitch and other vocal characteristics out of alignment with the Model so that the Pre-task may serve as a baseline from which alignment during the puzzle task could be determined.

An audiometer (AMBCO Model 650 A) was used to assess the level of hearing for each Speaker at the start of the experiment.

Two surveys were administered during the study. The first survey contained an adapted version of an abbreviated Big 5 Personality Domains inventory (Gosling et al., 2003) to assess where Speakers stood on these traits (see Appendix A). The results of the abridged version of the questionnaire are highly positively correlated with those of the full inventory (Gosling et al., 2003). This allowed for the Big 5 Personality Traits to be assessed given the time constraint of the experiment. Additionally, this first survey also asked Speakers to rate the friendliness and cooperativeness of their partner. The second survey asked Speakers about their language use and demographic information.

The computer software, Praat, was used for manipulating and playing the stimuli as well as recording the Speakers' responses.

Analysis. Fundamental frequency was isolated from the Speakers' speech recordings using Praat. Alignment in each condition was determined based upon whether the Speakers' pitch (Hz) during the Task was closer to the Model than during the Pre-task (Convergence) or further from the Model (Divergence). A non-significant difference in F0 between the Pre-task and Task calls for a Non-Shift on the part of the Speaker. The F0 from the Post-task recordings assessed whether alignment persisted after the Model-Speaker interaction ceased. A Repeated-Measures ANOVA looked for differences in F0 between groups (Convergence, Divergence, Neutral) and time-periods (Pre-task, Task, Post-task), as well as a possible interaction.

Phase II

A follow-up experiment utilizing perceptual (as opposed to acoustic) analysis was conducted to present a more holistic picture of alignment in response to the experimental manipulation in Phase I. Miller and colleagues (2013) as well as Pardo and colleagues (2013) address the importance of including both types of analyses when assessing phonetic alignment. The researchers suggested that the sociolinguistic function of alignment relies upon a combination of speech factors perceived in unison which, in turn, leads to changes in speech (Miller et al., 2013). While fundamental frequency represents one dimension of alignment, it is neither the only one nor is it isolated in everyday interactions. There are a myriad speech dimensions, including, but not limited to intensity, voice onset time, and utterance length, which are simultaneously perceived and factor into alignment. A perceptual study is best able to capture the interaction between these speech characteristics. In sum, Phase I attempts to quantify a concrete change in speech (F0) due to the direction of alignment employed by a conversational partner. Phase II complements the former by examining alignment through a more holistic lens to determine whether we can perceive and respond to shifts in such a way that may have served a useful sociolinguistic function throughout our evolution.

Subjects. Phase II was conducted at the same New England college as Phase I with 16 undergraduates (mean age = 18.8, SD = 0.856, female = 6, male = 9, not declared = 1). The subjects in Phase II will be referred throughout as *Listeners*. All Listeners were naïve as they were neither participants in Phase I nor were they aware of the experimental manipulation in Phase I. Listeners met the same English fluency and hearing requirements as did Speakers in the prior experiment and were compensated in the same fashion.

Experimental Design. This phase of the study consists of a single condition which required Listeners to judge the similarity of various speech segments taken from Speakers in Phase I relative to the Model. All Listeners performed the same task and were presented the same stimuli, albeit in a different order. Listeners were naïve concerning the Phase I condition (i.e. Convergence, Divergence, Neutral) and time-period during the interaction (i.e. Pre-task, Task, Post-task) from which the stimuli originated. The independent variables were the manipulation to the Model voice in Phase I and the time-period during the interaction. The dependent variable was the degree of perceived similarity as assessed by the Listeners.

Materials. The Speakers' and Model's speech recordings from Phase I served as the stimuli for Phase II. There were 630 potential recordings taken from Phase I. Due to a technical error, 10 Speakers were missing the final token. This brought the number of auditory stimuli to 620. All captured tokens were included as stimuli, including those where Speakers had made mistakes (regardless of whether the Speaker corrected themselves or not) in the naming and ordering of colored-shapes. Other features such as long pauses between the color and shape were included as well. An additional five recordings taken from the Model's speech in Phase I brought the final tally to 625 unique auditory stimuli.

Due to a discrepancy in the Speakers' phrasing between the different time-periods., the speech tokens chosen for analysis consisted of isolated colored-shapes rather than the full phrase (i.e. "Next to the *color 1* *shape 1* is a *color 2* *shape 2*"). Some Speakers used this full phrasing as instructed in the video, whereas others simply read off the colored-shape pairs (e.g. “blue square – green heart”). In contrast, all Speakers utilized the instructed phrasing during the Task itself. The tokens consisted of isolated colored-shapes (e.g. blue-square).

Five tokens were chosen for analysis which the Speaker uttered during the Pre-task, Task, and Post-task. The specific colored-shapes were chosen because both conversational partners uttered them. This is necessary because Speakers’ alignment is based upon their convergence/divergence relative to the Model. In turn, it is important to measure Speakers’ alignment on phrases that they utter in common with the Model. Each of the five tokens were a unique color-shape combination. The tokens consisted of the same colored-shapes for the Pre-task, Task, and Post-task. They are as follows:

1. Orange Square
2. Purple Circle
3. Blue Star
4. Black Square
5. Yellow Circle

A program written in Python using the computer program Psychopy was used to administer the experiment. The stimuli were presented to Listeners via over-the-ear headphones (Sennheiser HD 280 Professional) at a consistent volume. Instructions and response cues were presented on a monitor in Times New Roman, 12-point font with white lettering on a gray background. Listeners responded by pressing keys on a keyboard.

The hearing test and language demographic survey for this experiment were the same as those used in Phase I.

Procedure. Listeners were greeted by the experimenter upon entering the laboratory and, after giving consent, filled out the same language and demographic survey as did Speakers in Phase I. The hearing screening followed and was conducted in the same manner as in Phase I. Only those who met the hearing requirements (same as in Phase I) proceeded in the study.

A program written in Python using the Psychopy platform was used to administer the experimental task. The task consists of a series of trials. Each trial presented the Listener with a trio of acoustic stimuli using an *AXB* paradigm (see Table 1).

Table 1. The AXB paradigm used to compare Speakers' speech to the Model.

A	X	B
Speaker utterance from one time-period.	Model utterance of the same token.	Speaker utterance from a different time-period.

In each trial, Listeners were initially presented with auditory stimulus *A* simultaneously with the visual label "First" displayed on the monitor, followed by a 200ms inter-stimulus interval (ISI). The *X* stimulus was presented next alongside a "Middle" on-screen label, which was then succeeded by another 200ms ISI. Finally, *B*, was presented with "Last" shown on the screen. The monitor then instructed Listeners to determine whether the first or last recording was most similar to the middle one. Listeners were instructed to make their choice based upon *overall* similarity. Once Listeners had decided, they inputted their response on the keyboard. An 'f' response indicated that the first stimulus was most similar to the middle, whereas a 'j' response indicated that the last stimulus was most similar to the middle. This concluded one trial. A 500ms inter-trial interval (ITI) succeeded the Listener's response, and the cycle began anew.

Listeners completed a total of 620 trials. The order in which stimuli were presented within a trial was randomized so that each time-period from the Phase I interaction was individually compared to both other time-periods. Additionally, the order in which trials were presented was randomized so that each comparison was made once and each remaining comparison had an equal chance of being presented next. After completing all the trials, Listeners were compensated for their time in the same manner as in Phase I.

Analysis. Alignment was assessed with a one-sample t-test comparing Listeners' responses relative to chance. If there were no difference between Pre-task, Task, and Post-task stimuli in terms of similarity to the Model's speech, then Listeners would be expected to choose *A* (regardless of whether it is from the Pre-task, Task, or Post-task) 50% of the time. In each Phase I condition, Speakers were said to Converge if Listeners indicated Speakers' Task utterances to be more similar than the Pre-task utterance to the Model at a rate significantly higher than chance level. In turn, Speakers were said to Diverge if Listeners' responses revealed their Pre-task utterance to be more similar than the Task utterance to the Model at a rate significantly higher than chance level. Alternatively, there was a Non-Shift if the difference between the Listeners' Pre-task versus Task responses was nonsignificant. Alignment was said to persist beyond the interaction in cases of Convergence if Listeners determine the Post-task to be more similar than the Pre-task to the Model at a rate significantly higher than would be expected by chance. Likewise, alignment persisted in cases of Divergence if Listeners assess the Pre-task to be more similar than the Post-task to the Model at a rate significantly greater than would be expected by chance.

Between group differences were assessed with a one-way ANOVA to determine whether perceptions of alignment were influenced by the manipulation to the Model voice. Additionally, correlations were calculated between the alignment as rated by Listeners and the Speakers' personality factors as well as Speakers' perceptions of the friendliness and cooperativeness of the Model. Listeners' assessment of alignment between different time-periods was also determined for Speakers in each group. Furthermore, the patterns of Listener's preference ratings were compared to the changes in fundamental frequency from Phase I to compare and contrast the acoustic and perceptual analyses.

Results

Acoustic.

There was a main effect of time-period ($F(2, 38) = 9.686$, $p < .001$, $\eta^2 = .338$) on Speakers' pitch (see Figure 5). Speakers exhibited higher F0 during the Task compared to the Pre-task ($p = .001$) and Post-task ($p = .001$). A post-hoc Tukey LSD analysis showed that Speakers only exhibited elevated pitch during the Task; there was no significant difference between Pre-task and Post-task F0 ($p = .325$). The means and standard deviations are presented below (see Table 2).

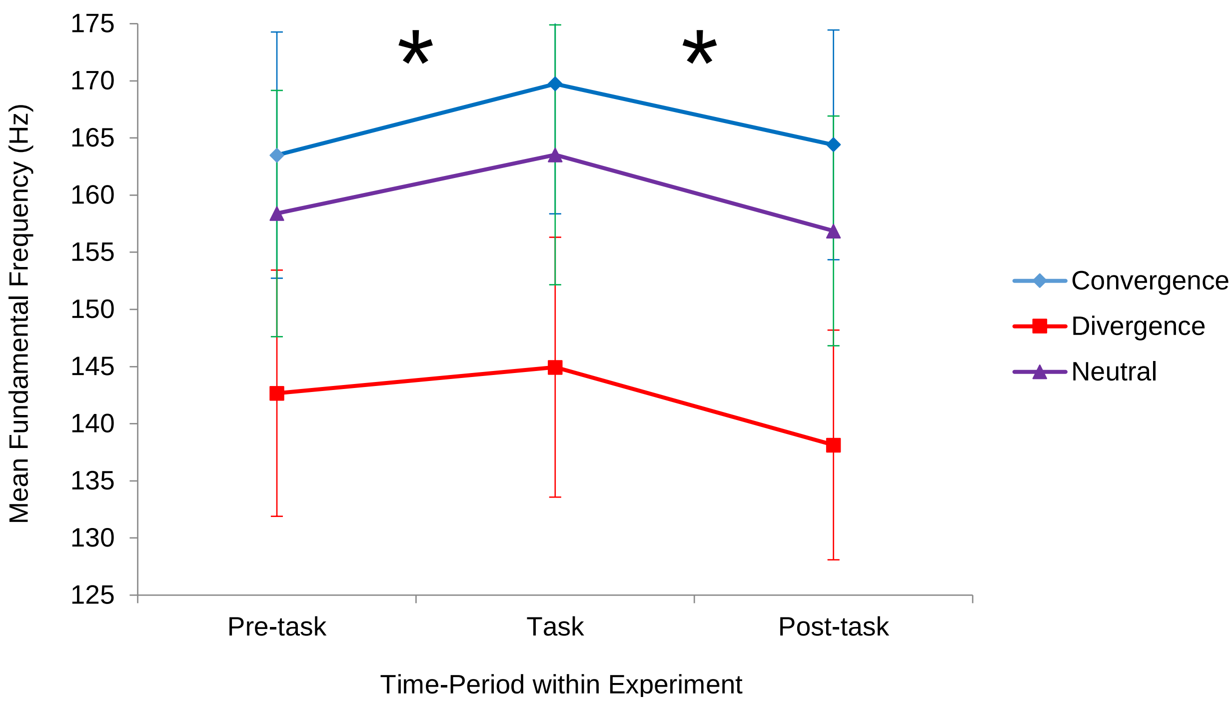


Figure 5. Speakers' fundamental frequency in each time-period for each condition. The Model's median pitch prior to manipulation was 93.4 Hz. Asterisks indicate significant differences between adjacent time-periods at an alpha level of .05.

A repeated measures ANOVA showed that there was no main effect of group ($F(2, 39) = 1.367$, $p = .267$, $\eta^2 = .065$) on F0 alignment (see Figure 5). The Convergence group did not differ from the Divergence ($p = .119$) group nor the Neutral ($p = .678$) group, and the latter two groups did not significantly differ from one another ($p = .247$). In other words, the different manipulations to the Model's pitch did not uniquely influence Speakers' F0 alignment behavior (see Figure 5). The means and standard deviations are reported below (see Table 2).

Table 2. Mean fundamental frequency values for Speakers in each condition (rows) and each time-period (columns) given in Hertz (standard deviations in parentheses). The mean pitch and standard deviations for all Speakers at each time-period are given in the bottom row.

Condition	Pre-task	Task	Post-task
Convergence	163.5019 (44.5232)	169.7334 (45.3862)	164.3979 (41.9861)
Divergence	142.6666 (40.3591)	144.9439 (44.6850)	138.1402 (37.8000)
Neutral	158.3880 (35.5283)	163.5281 (37.0362)	156.8714 (32.3934)
All	154.8522 (40.3207)	159.4018 (42.8348)	153.1365 (38.335)

In turn, there was no interaction effect ($F(4, 78) = .700$, $p = .700$, $\eta^2 = .027$) of condition and time-period on alignment (see Figure 5). The mean F0 values and standard deviations for Speakers in each condition and time-period are shown above (see Table 2).

Perceptual.

Listeners rated Speakers as significantly more similar to the Model in later time-periods than at earlier time-periods (e.g. Task is more similar than Pre-task to the Model) when each group was isolated (see Figure 6). When comparing Pre-task to Task, Listeners favored the latter significantly more often than chance ($t(671) = 2.489$, $p = .013$). For the Task to Post-task comparison, Listeners once again preferred the later time-period by a significant margin ($t(671) = 4.482$, $p < .001$). The same pattern was observed when the Pre-task and Post-task were compared ($t(671) = 3.148$, $p = .002$). Means and standard deviations for Listeners' ratings are presented below (see Table 3).

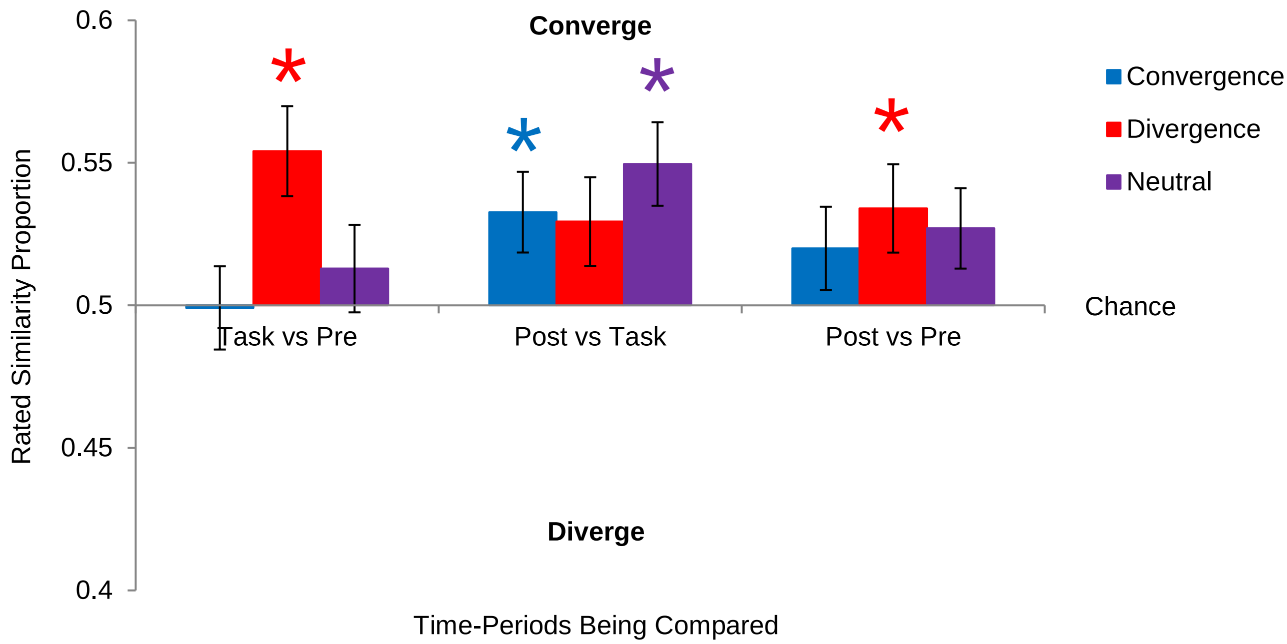


Figure 6. Listeners compare the relative similarity of Speakers in different conditions to the Model at different time-periods. Proportion represents how often Listeners picked the first listed (chronologically later) time-period of the two segments in each comparison. Asterisks indicate significant difference from chance at an alpha level of .05.

A one-sample t-test revealed that Listeners judged Speakers as trending towards greater similarity with respect to the Model over time in eight of the nine group/time-period intersections (see Figure 6), albeit only four were significantly different from chance (proportion of choosing later time-period > .5). These included Speakers in the Divergence group going from Pre-task to Task ($t(223) = 3.425$, $p = .001$) and Pre-task to Post-task ($t(223) = 2.183$, $p = .030$), as well as those in the Convergence ($t(223) = 2.520$, $p = .012$) and Neutral ($t(223) = 3.390$, $p = .001$) groups going from Task to Post-task. Again, Listeners in each case were rating Speakers as more similar to the Model in the later time-period than in the earlier time-period. An additional two intersections approach, but do not quite reach, statistical significance, namely Speakers in the Divergence group going from Task to Post-task ($t(223) = 1.894$, $p = .060$) and the Neutral group going from Pre-task to Post-task ($t(223) = 1.901$, $p = .059$). On the other hand, Listeners did not prefer a time-period for Speakers in the Convergence group going from Pre-task to Post-task ($t(223) = 1.346$, $p = .180$) and those in the Neutral group going from Pre-task to Task ($t(223) = .837$, $p = .404$). Only one group/time-period intersection—Speakers in the Convergence group going from Pre-task to Task—was judged to be more similar in the former time-period as opposed to the latter, however, this trend was non-significant ($t(223) = -.061$, $p = .951$). Means and standard deviations are reported below (See Table 3).

Table 3. Mean Listeners' ratings of similarity for Speakers in each condition at each time-period (standard deviations in parentheses). Values represent proportion of Listeners who rated the first-listed (chronologically later) time-period in each comparison as more similar to the Model. Asterisks indicate significant difference from chance (.5) at an alpha level of .05.

Condition	Task vs Pre	Post vs Task	Post vs Pre	**Mean for Group**
Convergence	0.4991 (.218)	*0.5327 (.212)	0.5200 (.218)	0.5172 (.216)
Divergence	*0.5541 (.236)	0.5294 (.232)	*0.5340 (.232)	0.5391 (.233)
Neutral	0.5129 (.229)	*0.5496 (.219)	0.5270 (.210)	0.5298 (.217)
Mean for Time-Period	0.5220 (.229)	0.5383 (.221).	0.5270 (.220)	0.5287 (.223)

A one-way ANOVA conducted on each type of perceptual comparison revealed a significant difference in similarity ratings across experimental groups in the Pre-task to Task comparison ($F(2, 669) = 3.514$, $p = .030$, $\eta^2 = .010$), but not in the Task to Post-task ($F(2, 669) = .489$, $p = .613$, $\eta^2 = .001$) nor Pre-task to Post-task ($F(2, 669) = .234$, $p = .791$, $\eta^2 = .001$) comparisons. Post-hoc analysis of the significant group difference showed stronger similarity differentiation in the Divergence group than in the Convergence group ($p = .030$), with the Neutral group in between. In other words, Listeners rated Speakers as increasing in similarity to the Model going from Pre-task to Task for the Divergence group more so than the Convergence group. The Neutral group did not differ from either the Convergence group ($p = .800$) nor the Divergence group ($p = .137$). The means and standard deviations for each condition/time-period intersection are shown above (see Table 3). Combined Listeners' ratings of individual Speakers for each condition are reported below (see Tables 4).

Table 4. Combined Listeners' ratings of individual Speaker behavior within each group. Each row within each condition represents Listeners' preferred time-period (most similar to Model) for one Speaker. Preferred time-period is listed in each cell (alpha level is .05).

Convergence			Divergence			Neutral		
Pre vs Task	Task vs Post	Post vs Pre	Pre vs Task	Task vs Post	Post vs Pre	Pre vs Task	Task vs Post	Post vs Pre
N.S.	Post	Post	Task	N.S.	N.S.	N.S.	N.S.	N.S.
Pre	Post	N.S.	Task	Post	Post	N.S.	Post	Post
Pre	Post	N.S.	Task	N.S.	Post	N.S.	N.S.	Pre
N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	Task	N.S.	Post
N.S.	N.S.	N.S.	Task	Task	Pre	N.S.	N.S.	N.S.
Task	Task	N.S.	N.S.	Task	Pre	N.S.	N.S.	Post
N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	Post
N.S.	N.S.	N.S.	Task	N.S.	N.S.	Task	N.S.	N.S.
Pre	N.S.	N.S.	N.S.	Post	N.S.	Pre	N.S.	N.S.
N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.	Pre
Task	N.S.	Post	Pre	N.S.	N.S.	N.S.	Post	N.S.
N.S.	Task	N.S.	Task	N.S.	Post	N.S.	N.S.	N.S.
Task	N.S.	Post	N.S.	N.S.	N.S.	N.S.	N.S.	N.S.
N.S.	N.S.	Pre	Pre	Post	N.S.	N.S.	Post	N.S.

Correlations.

Table 5. Correlations between time-periods and the personality factors. Pearson correlation in each cell (p-values in parentheses). A negative value indicates that the personality factor is significantly associated with less similarity between the comparison time-periods. Asterisks indicate significance at an alpha level of .05 (*) and .01 (**).

	Extroversion	Agreeableness	Conscientiousness	Neuroticism	Openness
Task vs Pre	.106 (.506)	-.066 (.678)	-.073 (.647)	-.063 (.693)	-.034 (.828)
Post vs Task	-.242 (.122)	-.179 (.258)	.276 (.077)	.154 (.331)	-.381 (.013)*
Post vs Pre	-.251 (.145)	-.251 (.109)	.220 (.162)	.111 (.482)	-.466 (.002)**

The Big Five Personality Traits (adapted from Gosling et al., 2003) were assessed for correlations with Listeners' ratings of Speakers (see Table 5). Openness shared a significant association with alignment. Specifically, Openness predicted a decrease in Speakers' perceived similarity relative to the Model between the Pre-task and Post-task (r = -.466, p = .002) as well as between the Task and Post-task (r = -.381, p = .013). There was no significant relationship between Openness and the interval between Pre-task and Task (r = -.034, p = .828). None of the other personality measures—conscientiousness, extroversion, agreeableness, and neuroticism—correlated with alignment as judged by Listeners. The Pearson coefficient correlations and associated p-values are reported above (see Table 5). Speakers' ratings of Model friendliness and cooperativeness were not significantly associated with alignment between any two time-periods (see Table 6).

Table 6. Speakers' mean ratings of Model friendliness and cooperativeness on a 5-point Likert scale along with standard deviations. Correlations with alignment for each time-period are depicted (p-values are listed in parentheses). The N/A for the correlations between Cooperativeness and time-period alignment are due to all Speakers assigning the same rating to the Model for this category.

	Mean	SD	Task vs Pre	Post vs Task	Post vs Pre
Friendliness	4.845	0.4999	.070 (.659)	-.252 (.108)	.009 (.954)
Cooperativeness	5.000	0.0000	N/A (n/a)	N/A (n/a)	N/A (n/a)

Alignment between two time-periods also predicted alignment with the third (see Table 7). Increased similarity, as rated by Listeners, between the Pre-task and Task was significantly associated with decreased similarity between the Task and Post-task (r = -.342, p = .027). On the contrary, the relationship between the Pre-task to Task comparison and the Pre-task and Post-task comparison approached, but did not quite reach, statistical significance (r = .279, p = .074). The final time-period relationship—Pre-task and Post-task compared with Task and Post-task—yielded a strong positive correlation (r = .466, p = .002).

Table 7. Correlations between time-periods indicate that similarity between any two time-periods predicts perceived alignment with the third time-period. Pearson correlation coefficient in each cell (p-values are listed in parentheses). Positive r values indicate that an increase in similarity between one set of time-periods predicts an increase in similarity between the comparison time-periods. Asterisks indicate significance at an alpha level of .05 (*), and .01 (**).

	Task vs Pre	Post vs Task	Post vs Pre
Task vs Pre	X	-.342 (.027)*	.279 (.074)
Post vs Task	-.342 (.027)*	X	.466 (.002)**
Post vs Pre	.279 (.074)	.466 (.002)**	X

Discussion

The goals of the present study were to determine whether alignment on the part of one partner influences the alignment of an interlocutor, to examine the similarities and differences between acoustic and perceptual assessments of alignment, and to identify any factors which serve as reliable predictors of vocal alignment behavior.

Aim #1: How does one partner's alignment influence the alignment of an interlocutor?

Acoustic. The increased fundamental frequency of Speakers in all three conditions during the Task relative to the Pre-task leads to multiple possible interpretations. One possibility is that this behavior reflects divergence on the part of Speakers, considering that the median Model frequency—as well as the adjusted Model F0 in the Convergence and Divergence conditions—are all below the mean Speakers' pitch for each of the three conditions (see Figure 5). In turn, an increase in pitch on the part of the Speakers would be deviating from that of the Model, a diverging behavior. This finding contradicts the general trend found in the literature, which suggests interlocutors' F0 converges throughout the interaction with a partner (e.g. Babel & Bulatov, 2011; Gregory & Webster, 1996; Gijssels et al., 2016). If this interpretation is correct, then the Speakers' behavior during the transition between Task and Post-task would represent a return to baseline immediately upon termination of the interaction. This is supported by the lack of a significant difference between Speakers' Pre-task and Post-task utterances. Prior research has also shown the immediate return to baseline in interaction-based alignment once the interaction ceases (Gijssels et al., 2016).

Another possible interpretation is that Speakers were changing their F0 for reasons other than alignment. This may make sense considering the lacking main effect of group (see Figure 5). Perhaps the methodological differences between the Task and the other two time-periods is sufficient to warrant a change in fundamental frequency. The unique features of the Task, while necessary for the experiment, may allow for such differences in vocal responses. Only in the Task do Speakers believe that they are participating in a live-interaction with a conversational partner (Model). The Task requires Speakers to participate in dialogue whereas the other two time-periods merely consist of monologue as they read off colored-shape pairs from a monitor. It is possible that these differences contribute to the increased F0 observed during the Task. The non-significant difference between the Pre-task and Post-task would make sense considering their fundamentally similar nature.

Yet, there is good reason to believe that alignment did occur. The perceptual results provide support that the difference between the Model's behavior between groups was detectable. The detection of the manipulation manifested in the difference between groups in the Pre-task to Task perceptual comparison (see Figure 6). If it were true that no alignment took place, then it must be the case that either a partner's F0 is not critical to changes in Speakers' pitch or that the manipulation was not salient to Speakers during the interaction. The former is unlikely to be true, considering the ample literature supporting F0's role in alignment and its unique presence in one partner's voice leading to greater imitation in the other partner (Babel & Bulatov, 2011). As for the latter claim, the lack of a main group effect does support the notion that the manipulation itself, or differences between the three conditions, may not have been perceived by Speakers. Perhaps the 10% shift articulated by the Model in the Convergence and Divergence conditions was not sufficient to warrant alignment. This is unlikely, considering that prior research has shown that participants' normal alignment behavior rarely exceeds a change greater than a .10 proportion relative to their partner (Babel & Bulatov, 2011). Additionally, a five percent shift has been shown to be sufficient to induce alignment (Gijssels et al., 2016).

One other possibility is that F0 alignment, if such behavior is truly absent, only did not occur because of the Model's atypically low F0. The Model's median pitch prior to any manipulation was 93.4 Hz. This frequency is much lower than most people produce; women typically have a range of 150-300 Hz, whereas men usually fall in the 75-150 Hz range (Weusthoff et al., 2013). The present study includes Speakers of both sexes. The extremely low

F0 of the Model may have made it difficult for Speakers to converge towards his pitch if it were outside of their normal range. Future research should use a Model with a higher F0 which resides in the middle of the typical F0 range to see if there is a group effect on the alignment of pitch.

Perceptual. Listeners' perception of Speakers as becoming more similar to the Model over time suggests converging behavior (see Figure 6). This behavior is consistent with the existing literature on alignment towards a Model when assessed through a perceptual lens (e.g. Miller et al., 2013). The perceptual measure employed in this experiment required that Listeners make their judgements based upon overall similarity, suggesting that there was convergence when all speech factors were assessed together in a holistic fashion.

The significant difference between groups in the Pre-task to Task perceptual comparisons suggests that there was greater convergence amongst Speakers when the Model Diverged as opposed to Converged (see Figure 6). This difference between Speakers' behavior in Convergence and Divergence conditions was expected, albeit in the opposite direction. Gregory and Webster (1996) showed that lower-status dyad members accommodated towards their higher-status counterparts while in conversation. Perhaps the perception of the Model's status was perceived to be higher amongst Speakers in the Divergence condition than in the Convergence condition. If this were the case, Speakers' convergence towards a high-status Model may have a functional purpose of decreasing social distance between the interlocutors (Giles et al., 1991). Perhaps the old proverb is true after all: we want what we can't have.

Another interesting component of the perceptual results concerns the alignment at each intersection of condition and time-period. While only four of the nine intersections were significantly different from chance, a total of eight leaned in the direction of convergence. This may hint at the arbitrary nature of the .05 cutoff value for significance. Some of these trends towards convergence may become statistically significant differences with more data. Future research should focus on recruiting additional Listeners for the second phase of the study.

Aim #2: Compare and Contrast the Acoustic and Perceptual Analyses.

The present study reveals a discrepancy in alignment between the two types of analyses! There is a potential for divergence in F0 throughout the conversation when examining alignment through an acoustic lens, however, an increase in perceptual similarity reflects holistic convergence during this same interaction. Regardless of whether Speakers truly diverged from the Model in F0 or if the increased pitch was simply due to the nature of the Task itself, Listeners judged the Speakers to mimic the Model *despite* the Speakers' deviation in pitch. In turn, there must be some other speech factor(s) which are more salient than F0 to Listeners as they make their similarity judgements. These potential vocal features include, but are not limited to, intensity and rate of speech.

If it is true that the Speakers' increase in fundamental frequency is the result of diverging behavior, then the findings from this study challenge Pickering and Garrod's (2004) Interactive Alignment Model. The common priming mechanism model suggests that alignment for all components of speech occurs through the same process which, in turn, should translate into the same direction of alignment. But this study finds opposing directions of alignment between F0 (acoustic divergence) and other speech factors (perceptual convergence). Gijssels and colleagues (2016) proposed that continuous and discrete levels of speech may have separate mechanisms of alignment. The researchers developed their theory based upon their data which suggested that pitch alignment does not adhere to potential priming criteria such as increased intensity with greater exposure and persistence beyond the termination of the interaction (Gijssels et al., 2016). The findings from the present study offers further support for separate mechanisms of alignment.

Differences in the speech factors incorporated in the acoustic and perceptual analyses warrant the inclusion of both when assessing vocal alignment (Miller et al., 2013; Pardo et al., 2013). This multi-faceted experimental design is supported by the dichotomous findings in the present study. While the acoustic analysis provides a concrete means by which to measure alignment, the perceptual lens reveals a more holistic picture that is more akin to the experiences of people in everyday conversation. Future research should look to isolate the individual factors of speech to develop a hierarchy of vocal characteristics which people use to assess alignment. Special attention should be paid to the relative importance of discrete and continuous traits.

Aim #3: Predictors of Alignment.

Greater perceived convergence between any two time-periods predicted the same relationship with the third in two of the three comparisons; one of these relationships was significant and the other approached, but narrowly missed, statistical significance (see Table 9). The two positive relationships manifested for the comparisons of Pre-task to Task and Pre-task to Post-task as well as Pre-task to Post-task and Task to Post-task. This suggests that either Listeners' perceived an increase in convergence throughout the experiment from Pre-task to Task to Post-task or that Speakers in the Pre-task sounded substantially more different than they did in the other two time-periods compared to the Model. As for the comparison between Pre-task to Task and Task to Post-task, greater perceived convergence with the former predicted less perceived convergence with the latter. This negative association would emulate the return to baseline effect observed in prior research (Gijssels et al., 2016).

Neither of the measures concerning Speakers' perception of the Model—friendliness and cooperativeness—were reliable predictors of Listeners' ratings of Speakers' alignment (see Table 8). This finding is surprising considering the literature linking increased convergence to positive perceptions of an interlocutor (Babel, 2010). The contradiction between the findings of the present study and prior research may stem from how perception of a partner is defined. Babel (2010) found the correlation exists between greater convergence and positive social biases of the interlocutor's nationality as measured by an Implicit Association Task. Here, the perceptions of the Model include friendliness and cooperativeness, and these characteristics are rated by Speakers on a Likert scale following completion of the interaction. The lack of a significant relationship in the present study may also be ascribed to Speakers' consistently high ratings of the Model's friendliness and cooperativeness. There were few ratings below the maximum score of five for friendliness, and cooperativeness only received ratings of the maximum score of five. This makes it is difficult to establish a firm relationship between perception of the Model and alignment. Perhaps a sample where Speakers' ratings of perceived friendliness and cooperativeness consists of a more Gaussian distribution would yield the sought-after relationship. This could be accomplished by either including more Speakers or multiple Models who differ in these characteristics.

The sole personality factor found to predict perceived alignment was Openness to experience (see Table 7). It is intriguing that prior research also identifies Openness as the lone Big Five personality factor associated with alignment behavior, however, the relationship was found in the opposite direction (Yu et al., 2013). Yu and colleagues (2013) found that greater Openness predicted an increase in converging behavior, rather than a decrease as shown in the current study. The apparent contradiction may in part be attributed to the operationalization of alignment. Yu and colleagues (2013) employ Voice Onset Time as their measurement, whereas a holistic measurement via perceptual judgements is used to identify a relationship with the personality factors in the current study.

Interesting relationships between Openness and perception are not limited to vocal alignment. A study conducted by Antinori and colleagues (2017) asserted that people who exhibit higher Openness are more susceptible to the mixing of two different visual stimuli. In turn, the researchers concluded that more Open participants have a more enhanced perception of visual stimuli relative to their less Open counterparts (Antinori et al., 2017). A difference in perceptual abilities between more and less Open people may explain the differences in their perceived vocal alignment. Perhaps Speakers with greater Openness notice auditory cues which are inaccessible to those with less Openness. This could lead to a difference in Speakers' own vocalizations which, in turn, can be parsed apart by Listeners. Alternatively, it may be the variance in Openness for the Listeners themselves which may make them more attuned to differences in Speakers' alignment behavior.

Future research should work to uncover the mediating role of Openness in alignment. Examining the personality factor along with specific speech factors other than fundamental frequency may shed more light on whether Openness is a facilitator of convergence or divergence. Assessing the degree of Openness amongst Listeners in addition to Speakers can parse apart the effects of the personality factor upon produced and perceived speech.

Conclusions

Alignment is a complex multi-faceted phenomenon which is by no means fully understood. The present study utilized a cooperative puzzle building activity which facilitates communication to examine alignment through both acoustic and perceptual lenses. A shift in one conversational partner's pitch revealed that the interlocutor's individual acoustic factors may differ in their direction of alignment within the same utterance despite a perceived uniform shift of either convergence or divergence. Further research is needed to identify a potential hierarchical relationship of speech factors that we use to assess alignment as well as the role of Openness to experience in mediating this relationship. There undoubtedly exists a viable link between alignment in pitch, speech, and general behavior.

Acknowledgements

I thank Professor Elizabeth Casserly for assisting me with planning, executing, and analyzing my study. Also, I would like to thank Professor Robert Outten for providing insight concerning the personality factors and Professor Michael Grubb for helping me with coding. Thank you to Olivia Wang, Rachel Scheub, and Annelise Pedro for helping me to pilot my experiments.

References

Antinori, A., Carter, O. L., & Smillie, L. D. (2017). Seeing it both ways: Openness to experience and binocular rivalry suppression. *Journal of Research in Personality, 68*(1), 15-22. doi:10.1016/j.jrp.2017.03.005

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society, 39*(04), 437-456. doi:10.1017/s0047404510000400

Babel, M., & Bulatov, D. (2011). The role of fundamental frequency in phonetic accommodation. *Language and Speech, 55*(2), 231-248. doi:10.1177/0023830911417695

Behroozmand, R., Korzyukov, O., Sattler, L., & Larson, C. R. (2012). Opposing and following vocal responses to pitch-shifted auditory feedback: Evidence for different mechanisms of voice pitch control. *The Journal of the Acoustical Society of America,132*(4), 2468-2477. doi:10.1121/1.4746984

Bilous, F. R., & Krauss, R. M. (1988). Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads. *Language & Communication, 8*(3-4), 183-194. doi:10.1016/0271-5309(88)90016-X

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. doi:10.1037//0022-3514.76.6.893

Chen, S. H., Liu, H., Xu, Y., & Larson, C. R. (2007). Voice *F0* responses to pitch-shifted voice feedback during English speech. *The Journal of the Acoustical Society of America, 121*(2), 1157. doi:10.1121/1.2404624

Cooper, M., & Yanagihara, N. (1971). A study of the basal pitch level variations found in the normal speaking voices of males and females. *Journal of Communication Disorders, 3*(4), 261-266. doi:10.1016/0021-9924(71)90032-3

Dias, J. W., & Rosenblum, L. D. (2011). Visual influences on interactive speech alignment. *Perception, 40*(12), 1457-1466. doi:10.1068/p7071

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: Quantifying the benefits of linguistic coordination. *Psychological Science, 23*(8), 931-939. doi:10.1177/0956797612436816

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences, 8*(1), 8-11. doi:10.1016/j.tics.2003.10.016

Gijssels, T., Casasanto, L. S., Jasmin, K., Hagoort, P., & Casasanto, D. (2016). Speech Accommodation Without Priming: The Case of Pitch. *Discourse Processes, 53*(4), 233-251. doi:10.1080/0163853x.2015.1023965

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, N. Coupland, H. Giles (Ed), J.

Coupland (Ed) & N. Coupland (Ed) (Eds.), (pp. 1-68). New York, NY, US; Paris, France: Cambridge University Press; Editions de la Maison des Sciences de l'Homme. doi:10.1017/CBO9780511663673.001

Giles, H., Taylor, D. M., & Bourhis, R. (1973). Towards a theory of interpersonal accommodation through language: Some Canadian data. *Language in Society, 2*(2), 177-192. doi:10.1017/S0047404500000701

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*(6), 504-528. doi:10.1016/s0092-6566(03)00046-1

Gregory, S. W. J., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of Personality and Social Psychology, 70*(6), 1231-1240. doi:10.1037/0022-3514.70.6.1231

Himberg, T., Hirvenkari, L., Mandel, A., & Hari, R. (2015). Word-by-word entrainment of speech rhythm during joint story building. *Frontiers in Psychology, 6*:727. doi:10.3389/fpsyg.2015.00797

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *Journal of the Acoustical Society of America, 108*(3), 1246-1251. doi:10.1121/1.1288414

Kaschak, M. P., Kutta, T. J., & Schatschneider, C. (2011). Long-term cumulative structural priming persists for (at least) one week. *Memory & Cognition, 39*(3), 381-388. doi:10.3758/s13421-010-0042-3

Lee, E. (2010). The more humanlike, the better? How speech type and users' cognitive style affect social responses to computers. *Computers in Human Behavior, 26*(4), 665-672. doi:10.1016/j.chb.2010.01.003

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*(6), 419-426. doi:10.1016/j.evolhumbehav.2013.08.001

Miller, R. M., Sanchez, K., & Rosenblum, L. D. (2013). Is speech alignment to talkers or Tasks? *Attention, Perception, & Psychophysics, 75*(8), 1817-1826. doi:10.3758/s13414-013-0517-y

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. doi:10.1177/026192702237958

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America, 119*(4), 2382-2393. doi:10.1121/1.2178720

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics, 40*(1), 190-197. doi:10.1016/j.wocn.2011.10.001

Pardo, J. S., Jordan, K., Mallari, R., Scanlon, C., & Lewandowski, E. (2013). Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory and Language, 69*(3), 183-195. doi:10.1016/j.jml.2013.06.002

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169-226. doi:10.1017/S0140525X04000056

Reich, C. M., Berman, J. S., Dale, R., & Levitt, H. M. (2014). Vocal synchrony in psychotherapy. *Journal of Social and Clinical Psychology*, *33*(5), 481-494. doi:10.1521/jscp.2014.33.5.481

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., & Schmidt, R. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science, 26*(6), 867-891. doi:10.1016/j.humov.2007.07.002

Weusthoff, S., Baucom, B. R., & Hahlweg, K. (2013). Fundamental frequency during couple conflict: An analysis of physiological, behavioral, and sex-linked information encoded in vocal expression. *Journal of Family Psychology*, *27*(2), 212-220. doi:10.1037/a0031887

Yu, A. C., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: subjective attitude, personality and "Autistic" traits. *PLOS ONE, 8*(9), 1-13. doi:10.1371/journal.pone.0074746

Appendix A

Psycholinguistics Survey

1. How friendly was your partner during the course of your interaction? Rate from 1 (very unfriendly) to 5 (very friendly).

2. How cooperative was your partner on a scale from 1 (very uncooperative) to 5 (very cooperative)?

3. Are you a member of a fraternity or sorority at Trinity College, or another campus?

4. Do you play a varsity sport for Trinity College or another school?

5. How many clubs are you a regular member of?

6. What is your class standing?

    a. Freshman
    b. Sophomore
    c. Junior
    d. Senior
    e. Other/non-traditional student

7. Are you a member of a culture house here at Trinity College (e.g., LVL, Hillel, I-house)?

8. On a scale of 1 to 5, indicate how well you have adjusted to campus life (1 = not adjusted at all; 5 = completely adjusted).

9. How well do you relate to other students on campus, or in other words, have you found your niche? Indicate on a 1 to 5 scale (1 = do not relate well; 5 = relate very well).

10. What category would best describe your major or intended major? Choose all that apply:

    a. STEM (science, technology, engineering, mathematics)
    b. Humanities (language, history)
    c. Art (visual or performing, music)
    d. Economics
    e. Political Science
    f. Interdisciplinary
    g. Other, please specify:

Rate the following statements on a scale from 1 to 7 in terms of how well they represent you.

Strongly Disagree	Moderately Disagree	Slightly Disagree	Neutral	Slightly Agree	Moderately Agree	Strongly Agree
1	2	3	4	5	6	7

11. I am extraverted and enthusiastic.

12. I am critical and quarrelsome.

13. I am dependable and self-disciplined.

14. I am anxious and easily upset.

15. I am open to new experiences and complex.

16. I am reserved and quiet.

17. I am sympathetic and warm.

18. I am disorganized and careless.

19. I am calm and emotionally stable.

20. I am conventional and uncreative.