COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 44 (2020) e12897

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12897

Natural Leaders: Some Interlocutors Elicit Greater Convergence Across Conversations and Across Characteristics

Uriel Cohen Priva,[a] Chelsea Sanker[b]

[a]*Department of Cognitive, Linguistic, and Psychological Sciences, Brown University*
[b]*Department of Linguistics, Yale University*

Received 23 March 2020; received in revised form 19 August 2020; accepted 25 August 2020

Abstract

Are there individual tendencies in convergence, such that some speakers consistently converge more than others? Similarly, are there natural "leaders," speakers with whom others converge more? Are such tendencies consistent across different linguistic characteristics? We use the Switchboard Corpus to perform a large-scale convergence study of speakers in multiple conversations with different interlocutors, across six linguistic characteristics. Because each speaker participated in several conversations, it is possible to look for individual differences in speakers' likelihood of converging and interlocutors' likelihood of eliciting convergence. We only find evidence for individual differences by interlocutor, not by speaker: There are natural leaders of convergence, who elicit more convergence than others across characteristics and across conversations. The lack of similar evidence for speakers who converge more than others suggests that social factors have a stronger effect in mediating convergence than putative individual tendencies in producing convergence, or that such tendencies are characteristic-specific.

Keywords: Convergence; Consistency; Corpus study; Individual differences; Interlocutor effects; Cross-characteristic; Social mediation

1. Introduction

Are there individuals who are more likely than others to change their performance such that it is more similar to their interlocutors? Are there interlocutors who are more likely than others to cause such shifts? Convergence, in which individuals' behaviors become more similar to their interlocutors, has been demonstrated in many

Correspondence should be sent to Uriel Cohen Priva, Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, 190 Thayer St., Providence, RI 02906. E-mail: urielc@gmail.com

Both authors contributed equally to this manuscript.

characteristics, both linguistic and nonlinguistic.[1] Many convergence studies find variation in degree of convergence across participants (e.g., Babel & Bulatov, 2011; Pardo, Cajori Jay, & Krauss, 2010). This variation is sometimes attributed to individual differences in convergent tendencies, which could indicate a broad cognitive trait producing these differences, but little work has tested whether individual tendencies are consistent across measurements of different linguistic characteristics or across tasks. The existence of individual differences in convergence is called into question by several studies that have looked for individual consistency but found weak evidence or no evidence (Pardo, Gibbons, Suppes, & Krauss, 2012; Pardo et al., 2018; Sanker, 2015; Weise & Levitan, 2018). Less work has addressed differences in the degree of convergence elicited by particular interlocutors. Consistency of convergence by speaker or by interlocutor across linguistic characteristics would suggest that social mediation of convergence is affecting different characteristics in the same way.

In this paper, we present a combined analysis of convergence in six linguistic characteristics, to probe factors that influence convergence: the characteristic measured, the speaker, the interlocutor, and the particular conversation. First, we present studies for convergence in six speech characteristics of different sorts, measured in the same conversational recordings: F0 median, F0 variance, speech rate, uh:um ratio, lexical information rate, and sentential conjunction. Even when the characteristics are z-transformed to put them all on the same scale, there is variability in the degree of convergence across different characteristics. Nevertheless, our subsequent cross-characteristic study finds consistency in convergence by interlocutor, indicating that there are external social factors at a high level that influence convergence in different linguistic characteristics similarly; some speakers elicit more convergence than others do. However, we do not find consistency by speaker, which suggests that there might not be any individual-specific cognitive trait which predisposes some speakers to converge more than others. Individual tendencies in convergence appear only within particular characteristics.

1.1. Individual differences in convergent traits

Degree of convergence may be influenced by traits of the participants, as some personal traits have been found to be significant predictors of degree of convergent behavior, such as higher openness and attentional focus scores (Yu, Abrego-Collier, & Sonderegger, 2013) or higher social desirability scores (Natale, 1975). Yu et al. (2013) regard this variation as evidence for individual differences in language processing. However, interpreting individual tendencies in convergence first depends on demonstrating that robust replicable individual tendencies exist; variation across participants within a task does not necessarily reflect inherent individual differences.

Individual differences have been identified in linguistic behaviors other than convergence, such as perceptual compensation for coarticulation (Yu, 2010), categorical perception (Kong & Edwards, 2016), lexical bias (Ishida, Samuel, & Arai, 2016; Stewart & Ota, 2008), and use of prosodic information to resolve syntactic ambiguity (Jun & Bishop, 2015). For a recent overview of existing literature on individual differences in linguistic perception, see Yu and Zellou (2019). It might be reasonable to expect similar individual differences in convergence, although convergence involves not just perception but also production and the link between them. Previous work has found inconsistent results on whether individual variation in perception and production is correlated: Some studies find a correlation between production of coarticulation and perceptual compensation for it (e.g., Beddor, Coetzee, Styler, McGowan, & Boland, 2018; Yu, 2019; Zellou, 2017) and between distinctiveness of vowels in perception and production (Perkell et al., 2004), while others do not find such correlations for coarticulation (e.g., Grosvald & Corina, 2012; Kataoka, 2011) or cue weighting (Schertz, Cho, Lotto, & Warner, 2015; Schultz, Francis, & Llanos, 2012). Some of the apparent perception–production correlations may reflect phonologized dialectal differences rather than individual differences (Harrington, Kleber, & Reubold, 2008). Failing to find a correlation does not necessarily mean that it does not exist. Schertz and Clare (2020) note that some of the varied results across studies may be due to differences in task design and analytical choices.

Most studies on individual tendencies in convergence have not retested individuals to establish their consistency, so interpretations are at risk of a fundamental attribution error. Many studies implicitly or explicitly assume that variation in convergence reflects traits of individuals, but only a few studies have tested the same individuals in multiple conversations or experimental tasks. There is some evidence that individuals have consistent tendencies in convergence in the same or very similar tasks, when measured in the same linguistic characteristic (Sanker, 2015; Wade, Lai, & Tamminga, n.d.), but these tendencies are much weaker across more dissimilar tasks such as shadowing and conversation (Pardo et al., 2018). In conversations, tendencies of an individual in producing convergence cannot be distinguished from tendencies of an individual to elicit convergence, which could confound the results when each speaker participates in a single conversation.

Is there a cognitive trait that differs across individuals, producing variation in convergence? Components of the Autism Quotient (AQ), in particular attention as mediated by social investment, provide a possible source of individual variation in convergence (Yu et al., 2013). In addition to social factors predicting convergence, direct manipulations of attention also show effects on convergence; speakers converge less when they have a larger cognitive load (Abel & Babel, 2017; Heath, 2017). Related to attention, differences in memory would be a possible candidate for individual variation, though Yu et al. (2013) found no relationship between working memory and convergence. Some work has identified cognitive traits or other individual attributes that correlate with perceptual behaviors, but it is unclear how such relationships would impact convergence. For example, neural differences in sensitivity to phonetic detail brain stem responses to speech stimuli, which have high test–retest reliability for individuals (Song, Nicol, & Kraus, 2011), correlate with accuracy of speech perception in noise and other phonological tasks (Hornickel, Skoe, Nicol, Zecker, & Kraus, 2009). Experience and learned knowledge also can influence perception: For example, larger social networks predict better perception in noise (Lev-Ari, 2018b), and speakers with larger lexicons rate nonce words with low phonotactic probability as more wordlike than speakers with smaller lexicons do (Large, Frisch, & Pisoni, 1998). Some work examines possible demographic characteristics that might align with differences in social and linguistic behavior. Gender is a frequently investigated demographic factor, but varied results in how gender relates to convergence make it difficult to formulate a clear analysis of what gender differences reflect, if they are capturing real differences. For example, Pardo (2006) found more convergence exhibited by men, but only for the speakers whose role in the task was to receive instructions, and Namy, Nygaard, and Sauerteig (2002) found more convergence among women, but only with one of the model talkers. Pardo, Urmanche, Wilman, and Wiener (2017) and Pardo et al. (2018) find larger effects of lexical factors on female participants than male participants.

If there is a broad cognitive trait driving individual variation in convergence, its effects should be consistent across different linguistic characteristics within which convergence is measured (Cohen Priva & Sanker, 2018; Weise & Levitan, 2018); an individual's pattern of convergence in one characteristic should be predictive of that speaker's convergence in other characteristics. Even if absolute convergence in each characteristic differs, the relative convergence in each characteristic should correlate across individuals. However, there is little work that compares convergence by each individual across characteristics and most studies that have looked for such patterns found no effect (Bilous & Krauss, 1988; Pardo et al., 2012; Sanker, 2015; Weise & Levitan, 2018). It is nonetheless possible that the null result in these studies can be attributed to lack of power rather than the absence of an effect.

Not all theories that account for convergence make the same predictions about individual consistency or consistency across different speech characteristics. Exemplar Theory and related episodic accounts of convergence propose that memories of incoming speech are stored as exemplars, which are aggregated into representational clouds; some exemplars may be given more weight than others, based on social factors and other influences. Subsequent speech production draws on these exemplar clouds, which results in greater similarity to recently heard speech (Goldinger, 1998; Pierrehumbert, 2002). The weighting of acoustic details in these representations can vary based on which characteristics are salient given the language and the task (Johnson, 1997). Speaker-specific variation in convergence could be integrated into the model merely by allowing individual differences in storage of particular details and weighting of input from different contexts, though individual differences are not inherently predicted by this theory. Episodic representations can also depend on phonological status (Mitterer & Ernestus, 2008), and they may relate to perceptual processes (Fowler, Brown, Sabadini, & Weihing, 2003), so the existence of individual differences in perception and phonological representations suggests the existence of individual differences in convergence. The theory is not constrained such that greater convergence in one characteristic would necessarily predict greater convergence in other characteristics, either when separated by speaker, by interlocutor, or by conversation. An individual could exhibit variation in how much each linguistic characteristic is influenced by input from interlocutors. Pardo et al. (2012) specifically address this in the discussion of the lack of individual consistency that they found in convergence across different characteristics, concluding that different phonetic characteristics were salient to different pairs of interlocutors.

Some theoretical accounts of convergence explicitly predict a correlation across different linguistic characteristics, based on alignment at one level facilitating alignment at other levels (Pickering & Garrod, 2004, pp. 174–175). This theoretical alignment might be observed in neural synchronization, which has been found to increase along with behavioral synchrony (Yun, Watanabe, & Shimojo, 2012). In contrast to the phonological mediation assumed in exemplar models, some theories of convergence propose a more automatic perception–production link: If incoming speech automatically activates motor plans that are subsequently reflected in production (e.g., Pickering & Garrod, 2013), effects should be consistent across tasks, and different phonetic characteristics should exhibit the same influences, though lexical and syntactic characteristics might behave differently. Pardo et al. (2018) argue that such a perception–production link cannot be automatic, because it varies based on context, resulting in different convergence behavior of the same speaker in different tasks.

Consistency of an individual in convergence does not mean that convergence will necessarily be the same in each linguistic characteristic, but rather that variation across speakers or across interlocutors will be parallel for different characteristics. Previous work clearly demonstrates that convergence can vary substantially depending on what characteristic has been used to measure it. In the same task, there can be substantial differences in overall convergence in different measures (e.g., Babel, 2012; Sanker, 2015) and also differences in effects of conditioning factors (e.g., Bilous & Krauss, 1988; Pardo et al., 2017). Differences in convergence across characteristics have not been a major focus of convergence research, but there are several possible explanations. One explanation for this variation is cumulative priming (Kaschak, Kutta, & Jones, 2011; Oben & Brône, 2016): Different constructions and forms are presented a different number of times, so the ones which were presented more have been primed more and will accordingly have a larger impact on speakers' subsequent productions. Some work has suggested that phonemes with more variable realizations are more likely to exhibit convergence (e.g., Babel, 2009, pp. 141–142), based on suggestions that convergence develops out of naturally occurring variation in speech (e.g., Delvaux & Soquet, 2007; Pierrehumbert, 2002). There also might be variation due to differences in the salience of each characteristic (Johnson, 1997; Pardo et al., 2012). There are also differences between results for acoustic measures of convergence and holistic measures using AXB tasks in which listeners make decisions about similarity: Acoustic measures are sometimes significant predictors of these holistic perceptual measures (e.g., Pardo et al., 2017), though other studies have found a lack of correlation between holistic measures and acoustic measures (e.g., Babel & Bulatov, 2011; Pardo et al., 2010).

1.2. Social mediation

Variation in degree of convergence might be due to social factors rather than or in addition to individual cognitive differences. Convergence studies are usually designed to account for possible variation by speaker, using multiple participants, but they often use a single model talker in shadowing tasks (e.g., Babel, 2010; Yu et al., 2013) or have each individual participated in a single conversation with one interlocutor (e.g., Abel & Babel, 2017; Pardo, 2006). However, some studies use multiple model talkers and find substantial variation, indicating that the interlocutor or model talker is an important factor to consider (Babel, McGuire, Walters, & Nicholls, 2014; Pardo et al., 2017). Certain individuals might elicit more or less convergence, either based on their behavior or based on what the listeners know or think they know about them. Many studies have found that aspects of the interlocutor or model talker's perceived identity or the speaker's perspective toward that identity can influence convergence. Factors that have been found to influence convergence include native language (Kim, Horton, & Bradlow, 2011), perceived standardness of the model talker's dialect (Weatherholtz, Campbell-Kibler, & Jaeger, 2014), interlocutor status (Gregory & Webster, 1996), and attitude toward a model talker (e.g., Babel, 2010; Yu et al., 2011) or interlocutor (Pardo et al., 2012; Sanker, 2015). The observed effects of these social factors on convergence are not always consistent across different studies.

Social mediation via the speaker's opinion of the interlocutor is a likely source of interlocutor effects in convergence. In studies that specifically manipulate the subject's opinion of the model talker, a more positive opinion predicted greater convergence (Bourhis & Giles, 1977; Yu et al., 2011). In conversations in which subjects were allowed to develop opinions naturally, liking of the interlocutor also predicted greater convergence (Pardo et al., 2012; Sanker, 2015). Greater perceived attractiveness of a voice is also predictive of how much convergence it elicits (Babel et al., 2014). Communication Accommodation Theory proposes that speakers use convergence and divergence to manipulate social distance from interlocutors (Giles, Coupland, & Coupland, 1991; Giles, Taylor, & Bourhis, 1973). However, socially mediated convergence does not necessarily need to be done consciously; interlocutor effects may result from greater weight given to input from speakers who are viewed more positively (Chartrand & Bargh, 1999), which predicts that there should be consistency across different linguistic characteristics in the convergence elicited by different interlocutors. If greater attention or engagement with some interlocutors or in some contexts increases convergence, this could also be consistent with stronger priming in goal-oriented tasks which require stricter understanding than conversations do (Reitter, Moore, & Keller, 2006); however, other work has not found this difference (e.g., Pardo et al., 2018). One possible mechanism for how certain input is given greater weight is offered by Jiang et al. (2015), based on interpersonal neural synchrony; the speakers whose partners synchronize more with them tend to be perceived as "leaders," based on a range of skills in communication and reasoning.

However, the social status meaning of "leader" is not implicit in our use of the term, which instead refers to the metaphorical leaders who drive the form of variable linguistic characteristics used within a conversation, that is, speakers who elicit higher levels of convergence than others do. While higher social status does predict greater convergence (e.g., Gregory & Webster, 1996), this is far from the only trait that is predictive of how much convergence a speaker elicits; some of the factors found to contribute to being a leader of convergence are discussed above. The term refers to the individuals who elicit the most convergence, but it is not meant to suggest that eliciting convergence is categorical. This terminology is borrowed from discussions of “leaders of sound change”; some work links phonetic convergence to sound change (e.g., Yu, 2013), which makes the connection particularly relevant. Some work on the actuation of sound change addresses the issue of whether sound changes tend to be led by certain individuals, and whether there are any consistent traits that characterize those individuals. Part of this question is the issue of who initiates change: At least for changes that are attributed to coarticulation, innovations may originate with speakers who compensate less for coarticulation (Yu, 2010; Yu & Lee, 2014). The other main aspect of this question is the issue of whose innovations are likely to be adopted by others: The individuals whose variants are likely to spread are those who are highly socially connected or otherwise influential (Baker, Archangeli, & Mielke, 2011; Labov, 2001).

Degree of convergence is in part predicted by aspects of the interaction and the relationship between the interlocutors (e.g., Bane, Graff, & Sonderegger, 2010; Gregory & Webster, 1996; Pardo et al., 2012). If variation is driven by situational or interactional factors that increase convergence via broad mechanisms such as increased attention, there should be consistency between participants within the same conversation and across different linguistic characteristics within that conversation. Little work has systematically tested whether the convergence exhibited by a conversing pair in one characteristic was predictive of their convergence in other characteristics within the same interaction. In a comparison across a small number of conversations, Sanker (2015) found only a weak trend toward positive correlations across characteristics. With a larger dataset, Weise and Levitan (2018) found no correlation in convergence across characteristics. Cohen Priva and Sanker (2018) found a correlation only between closely related characteristics: F0 median and F0 variability.

Some variation across interlocutors may be due to differences in amount of exposure to their speech. More exposure to a construction or a lexical item increases the probability that a subject will subsequently produce that item (Kaschak et al., 2011; Oben & Brône, 2016), though cumulative effects in phonetic convergence are less clear (e.g., Babel, 2012; Gijssels, Casasanto, Jasmin, Hagoort, & Casasanto, 2016). In conversational tasks, different amounts of speech produced by each individual would produce different levels of exposure and might result in different degrees of convergence; talkative individuals might elicit more convergence than others do. However, variation in the degree of convergence elicited by different interlocutors or model talkers has been found even in tasks with equal exposure to each voice (e.g., Hwang & Chun, 2018; Pardo et al., 2017), indicating that there must be factors other than just exposure.

Many studies include each individual only in a single interaction, and thus cannot separate out the individual contributions of each speaker to convergence in a conversation, so tendencies by speaker, by interlocutor, or by conversation would all produce the same pattern of variability. A study that includes each individual in multiple interactions with different partners makes it possible to separate speaker, interlocutor, and conversation as possible factors with distinct tendencies in convergence.

2. Methods overview

2.1. Corpus

The data for this study are the Switchboard Corpus (Godfrey & Holliman, 1997), a large collection of telephone conversations. Each speaker was randomly paired with other speakers and given a topic for each conversation, providing a large corpus of natural speech data for many speakers in similar conversations with several different partners. Recordings include associated speaker identification data that can be used to compare all instances of that speaker in different conversations. Each side of the conversation is a distinct recording, so measurements can reliably be taken for each speaker separately.

Each conversation has associated information quantifying the clarity of the recording. To ensure reliable acoustic measurements (F0 median and variance), calls were omitted if they had high levels of background noise, echoing, or other issues, as indicated in their clarity rating. This resulted in a total of 464 speakers in the data used for acoustic characteristics. For measurements that did not depend on acoustic form, no conversations were omitted, so these measurements were based on 518 speakers. Conversations took 6:20 min on average (the median was 5 min). Word boundaries were based on the manually corrected word annotations produced at MS State (Harkins, Feinstein, Lindsey, Martin, & Winter, 2003). The word annotations allow measurement of word duration.

Previous work has used similar data from pairing individuals with strangers for conversations (e.g., Natale, 1975; Pardo, 2006) or presenting listeners with unfamiliar voices in shadowing tasks (e.g., Pardo et al., 2017; Yu et al., 2013). Work arguing for individual tendencies in convergence uses such data, based on by-participant variation correlating with traits such as higher openness and attentional focus scores (Yu et al., 2013) or higher social desirability scores (Natale, 1975). However, the real test of individual tendencies in convergence is in behavior across different conversations or tasks. If individual variation exists, it should be consistent across conversations even if there are other factors also contributing to variation. If individual tendencies in convergence can only be found within replications of the same task, it is unclear whether the consistency is due to something inherent about the individuals themselves, or something about the particular combination of task, speaker, and interlocutor or model talker.

2.2. Measuring convergence

There are several different approaches to measuring convergence. We adopt the method introduced by Cohen Priva, Edelist, and Gleason (2017) and Cohen Priva and Sanker (2018). The benefits of this method are discussed by Cohen Priva and Sanker (2019), but it has not yet been widely adopted, so we describe it here as well.

In shadowing tasks, speakers (S) are exposed to a pre-recorded reference value (R) and repeat after the recorded items produced by the model talker (e.g., Babel, 2012; Goldinger, 1998; Pardo et al., 2017). Speakers' productions of each linguistic characteristic are measured before the exposure (S_b) and after the exposure to the model talker (S_R). In conversational tasks, speakers' baseline values can be measured from the beginning of the conversation or in interactions with other interlocutors, as compared to their speech during or after the conversation with the interlocutor. Convergence is a change in which the speaker becomes more like the model talker, that is, a change from S_b to S_R that makes S_R more like R. Following this conceptualization, Cohen Priva et al. (2017) model convergence using linear combination in a regression model, as in (1). The degree of convergence is measured as β_R, the relative importance of the interlocutor in predicting the performance of speakers after the interaction, relative to their *consistency* (*self-correlation*), which is measured as β_{S_b}. Thus, in this model, speakers' performance is predicted by a combination of their self-consistency, the performance of their interlocutor, and noise. Cohen Priva and Sanker (2019) show that this method is not susceptible to artifacts caused by extreme initial values or the distance between the speaker and interlocutor, while measuring convergence with the commonly used difference-in-difference method produces artifacts of both. Earlier formulations of this model (Schweitzer & Lewandowski, 2013) used a coefficient just for the interlocutor and modeled self-consistency β_{S_b} *Sb* as a random intercept.

$$S_R \approx \beta_0 + \beta_{S_b} S_b + \beta_R R + \varepsilon. \tag{1}$$

In this study, convergence is measured in two stages. First, each linguistic characteristic (e.g., F0) in each conversation is measured and processed separately. For every characteristic, all of the measurements for each speaker in each conversation are summarized to a single statistic (the median, interquartile range, or average, depending on the characteristic—the statistic used for each characteristic is explained below). Thus, each conversation side is represented by a single value for each characteristic. This step is illustrated for speech rate value in the top three and bottom three boxes in Fig. 1.

Second, for each characteristic, each conversation side is summarized based on (a) the speaker's performance during the conversation (equivalent to S_R above), (b) the speaker's mean performance in other conversations, to be used as the speaker's baseline S_b, and (c) the interlocutor's mean performance in other conversations, to be used as the convergence target (R).

Finally, these three data points for each characteristic are used to predict convergence, as illustrated for speech rate in Fig. 2.

The method measures convergence to the interlocutor's baseline, rather than to the interlocutor's performance in the shared conversation. This way of measuring convergence is necessary in order to exclude conversation-specific changes that are not convergence (aspects of the environment or conversation that influence both speakers similarly), and to avoid attempting to estimate how a speaker's performance is influenced the interlocutor's performance, when both influence each other.

Comparing a speaker's performance to the interlocutor's performance in the shared conversation could produce the appearance of convergence even in the absence of actual convergence, because the particular conversation might influence both interlocutors similarly. For example, both speakers might increase their F0 and speech rate because they are excited about a particular topic, in which case they are not converging even though

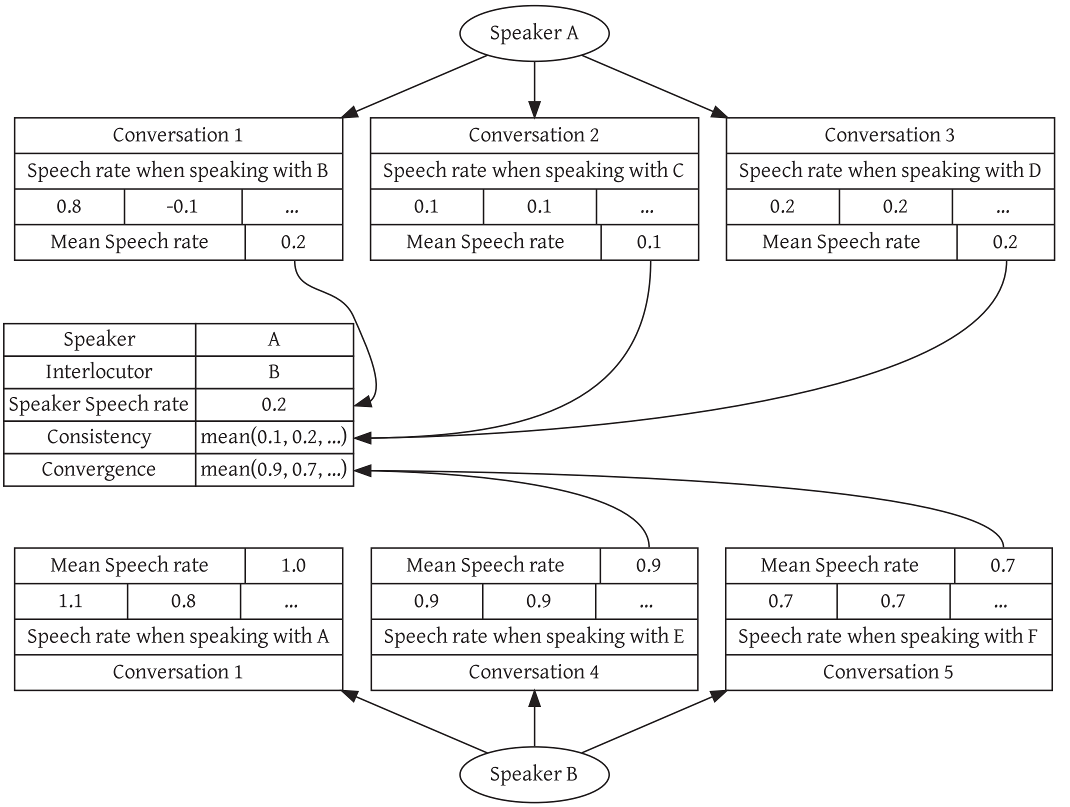

Fig. 1. An illustration of the first two steps in the procedure, as they apply to speech rate. First (top three and bottom three boxes), the mean of multiple speech rate values (log observed over expected) is taken, yielding a single point per conversation side. Second, for each conversation side, the speaker's performance provides the predicted value, the mean of the speaker's performance in other conversations is used as the speaker's baseline (S_b), and the interlocutor's performance in other conversations is averaged to yield the interlocutor's baseline (R).

their speech is becoming more similar. When the subject matter involves a particular typical speaking style (e.g., using low frequency words or slower speech rate), the speakers would seem to converge if they both adopt that speaking style, even if they ignore each other's performance. Such effects are likely to be responsible for the significantly greater similarity of an interlocutor to a speaker as compared within a shared conversation than as compared to that speaker's productions from other conversations (Gregory & Webster, 1996).

The other issue inherent in measuring a speaker's shift toward the interlocutor's productions in the shared conversations is that both speakers might be converging. If the interlocutor is converging, then the interlocutor's productions in the shared conversation are influenced by the speaker's performance in that conversation. It is undesirable to have a predictor (the interlocutor's productions in the shared conversation) that is known to be

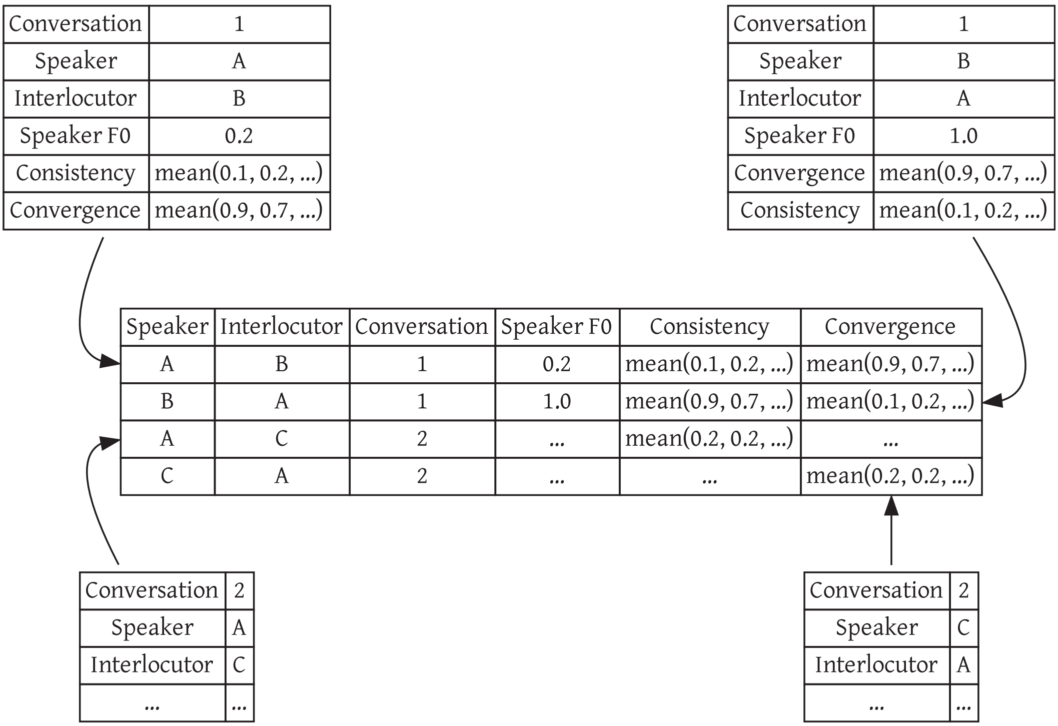

Fig. 2. An illustration of the third and final step in the procedure, in which individual data points are combined to create the collection of data points for the analysis. Every conversation provides two summary data points, one for each side of the conversation.

influenced by the dependent variable (the speaker's productions in the shared conversation). In addition to making measurements of convergence unreliable, this issue makes it difficult to separate out the contributions of each partner to convergence, which are crucial when looking for possible individual tendencies in converging or eliciting convergence.

A simple solution to both of these issues is to use a baseline for the speaker's productions and also for the interlocutor's productions. Speakers have high consistency in many linguistic characteristics, so their baselines from other conversations are a meaningful measure of the consistent characteristics of their speech, which will be present in the shared conversation. In the Switchboard corpus, in which speakers interact with multiple different interlocutors, a natural way to approximate speakers' baselines and their interlocutor's baselines is to average their performance from other contexts, in which they were interacting with other conversation partners. While this method of using baselines from other conversations may decrease how much convergence is captured, convergence is still apparent when tested in this way (Cohen Priva et al., 2017; Cohen Priva & Sanker, 2019). Establishing reliable baselines depends on having a large corpus, so that baselines are averaged across enough conversations to not be thrown off by any particular conversation; the median number of conversational partners for speakers in our dataset was 9 (Q1:4, Q3:12).

The use of multiple data points per speaker and interlocutor necessitates adding random effects structure for repeated measures. Beyond their necessity when using repeated measures, random effects can be used to examine the effect of individuals when that variation is itself a point of interest. For example, in our design, random slopes for β_R can be used to find individual differences in degree of convergence. The behavior of random slopes is similar to what we would expect from an interaction term in more traditional analyses, except at the group level. For instance, if some individuals converge more than others, or cause more convergence, a random slope could capture individual-level variation for the convergence coefficient. Adding a random slope does not offer a statistical test of whether a particular individual is more convergent than others, but model comparison between a model that contains a particular slope and a model that omits that slope should reveal whether individual-level variation in convergence exists. If variation by speaker or by interlocutor exists, then adding these random slopes will significantly improve the model, as reflected in model comparison.

In linguistic studies on individual differences, β values from random effects in regression models are sometimes used as per-participant measurements that can be tested for correlations with other linguistic characteristics, for example, when comparing behavior across multiple tasks (e.g., Pardo et al., 2018). Some studies instead fit separate models for each subject and use the coefficient from the fixed effects in each model as the per-subject measurement (e.g., Schultz et al., 2012; Yu, 2019); the latter method treats each individual independently and may overestimate individual differences, while using random slopes from a single model is shaped by the assumption that there are group-level effects which variation occurs around. Linguistics studies do not usually demonstrate the existence of individual differences by showing that such random effects significantly contribute to the model, but some do (e.g., Cohen Priva & Sanker, 2019; Drager & Hay, 2012). This method is seen more often in other fields. See Dingemanse, Kazem, Réale, and Wright (2010) and Martin, Nussey, Wilson, and Réale (2011) for discussion of using random slopes to detect individual differences in biology.

In the models we present, there are per-speaker and per-interlocutor random slopes for β_R.[2] The different random slopes for β_R measure distinct properties. Per-speaker slope for β_R attempts to capture by-speaker variance in convergence: whether some speakers converge more than others (likelihood to coverge). Per-interlocutor slope for β_R attempts to capture variance among interlocutors in eliciting convergence: whether some interlocutors are converged to more than others are (likelihood to lead convergence). These are distinct properties; we do not assume that those who converge more would necessarily elicit less convergence. We do not include random slopes for β_{S_b} (speakers' reliance on their own baseline), because this is not our variable of interest in the current investigation.[3] We follow Barr, Levy, Scheepers, and Tily (2013) in using all relevant random effects for our variable of interest, but we have a more parsimonious set of other random effects.[4]

2.3. Characteristics

Convergence was measured in six linguistic characteristics. The goal was to have a broad range of characteristics, both related and unrelated. Four of the six, F0 median, F0 variance (IQR), speech rate, and *uh*:*um* ratio, were examined in Switchboard by Cohen Priva and Sanker (2018). Two characteristics were added to expand the range of characteristics used and to include additional non-phonetic measures: lexical information rate and sentential conjunction. All six characteristics are described below.

F0 median F0 is the measure of the frequency of wave cycles produced by the vibration of the vocal folds. The measurement of frequency was converted into the mel scale, which provides a better approximation of human perception than Hz (Stevens, Volkmann, & Newman, 1937). Medians were used because F0 measurements of individual words are often noisy, with outliers that may be due to pitch tracking errors; the median is less influenced by outliers than other measurements of general tendencies are.

F0 range F0 range was measured as the log of the ratio of the 75th percentile to 25th percentile of F0 measurements in mels (IQR). As with F0 median, IQRs were used to measure variance due to the inherently noisy measurement of F0 in individual words: Many measures of variability are highly influenced by outliers, but IQR is not.

Speech rate Speech rate in a conversation was measured as the mean log speech rate of individual utterances. Following Cohen Priva et al. (2017), point-wise speech rate was measured as the actual utterance duration (including pauses) divided by the expected utterance duration. Expected utterance duration was calculated as the sum of the predicted durations of words in the utterance, each calculated as the predicted value of a linear regression using the median duration of that word in the entire corpus, the length of the utterance, and the distance from the end of the utterance. Unlike F0 measurements, speech rate was calculated based on hand-corrected values, so every measurement is reliable and the use of averages is appropriate.

***uh*:*um* ratio** This measure was calculated as the log odds of *uh* versus *um*, two frequently used filled pauses in English. The use of one or the other seems to be influenced by processing factors (e.g., Clark & Fox Tree, 2002) as well as other factors such as gender (Acton, 2011). Log odds were calculated as the predicted values plus the residuals of a logistic regression between the number of *uh* uses and *um* uses in each conversation side, which could be evaluated even when a subject never used one or the other. The use of log odds is standard for binary values.

Lexical information rate This measure was calculated following Cohen Priva (2017), as the mean negative log unigram predictability (the unigram entropy) of non-function words used in each conversation. Word counts were estimated using a combination of the Buckeye (Pitt et al., 2007), Fisher (Cieri, Graff, Kimball, Miller, & Walker, 2005; Cieri, Miller, & Walker, 2004), and Switchboard (Godfrey & Holliman, 1997) corpora. Lexical information rate was chosen as it is easy to measure and average across a conversation (as opposed to tracking the use of individual words), and because it has been linked with speech rate (Cohen Priva, 2017), which increases the odds of finding common behaviors between the two measures. Unigram lexical information rate captures the diversity and rarity of the words used by speakers. There are several ways that speakers may converge in lexical information rate. They may shift to more or less formal language registers or shift toward a shared specialized lexicon. To focus on such shifts, rather than on shifting choices in frequent words (e.g., expressions such as *you know*, *that's right*), we excluded the top 1% most frequent words.[5] Averages were used for the summary statistic because the corpus is large enough that the accuracy of each measurement is reliable (Cohen Priva & Jaeger, 2018), and also because this is the measure used in existing lexical rate studies.

Sentential conjunction This characteristic measured the use of *and* to connect sentences, as in (2). In these cases, *and* is not strictly required, and the meaning of the sentence would change little if *and* were omitted, so sentences with and without this initial conjunction can be compared. These sentence-initial conjunctions have been described as discourse markers; they make clear the continuity between a statement and the preceding discourse and have various contextual predictors (Dorgeloh, 2004; Schiffrin, 1986).

For each conversation side, we counted the number of times "*and* pron-nom" appeared, where pron-nom could be any of *I*, *she*, *he*, or *they* (~22,360 tokens). We excluded cases in which the preceding word was a given name (~15 tokens), as in (3), sequences containing *you and I* (~140 tokens), and word sequences of "*my* kinship-term *and I* " (~400 tokens), to exclude cases such as (4). Such constructions comprise the majority of "*and* pron-nom" sequences that are not cases of sentence conjunction. As with *uh*: *um* ratio, we used a logistic regression to calculate the log odds between pron-nom with and without a preceding *and*, providing a comparison of pronominal sentences with or without sentential conjunction. There were ~242,700 uses of pron-nom, of which ~21,800 were preceded by *and*. The use of log odds is standard for binary values.

(2) ... I certainly wouldn't object to it *and* I think random [testing for drugs] is probably, you know, the only really fair way ... (SW2638A)

(3) ... that *Dan and I* are going to ... (SW3323B)

(4) ... in the old days when *my wife and I* both worked ... (SW4238A)

As a preliminary step, we verified that all six characteristics exhibit convergence. We therefore built mixed effects linear models following the formulation in (1), with random intercepts for the speaker, the interlocutor, the conversation, and the topic of conversation (as listed in the Switchboard corpus), as well as per-speaker and per-interlocutor random slopes for β_R (looking for per-speaker variation in convergence and per-interlocutor variation in convergence, respectively). All of the predictors and predicted values were z-transformed across all data points to allow for easy comparison across the six characteristics. This yields the formula provided in Table 1 (in lme4 syntax).

Speakers' baselines (β_{S_b}) model speakers' self-consistency well, making per-speaker intercepts redundant. In five of the six characteristics, the model fitted no variance to a per-speaker intercept. In the one remaining case, the per-speaker intercepts were fitted

Table 1
lme4 formula used to fit single-characteristic convergence models

lme4 Syntax	Explanation
Speaker ~	Speaker's performance in a conversation (S_R)
1	Intercept, expected to be zero (β_0)
+ speaker.baseline	Speakers' baseline: captures consistency (β_{S_b} S_b)
+ interlocutor.baseline	Interlocutors' baseline: captures convergence (β_R R)
+ (1 | speaker)	Differences by speaker
+ (1 | interlocutor)	Differences by interlocutor
+ (1 | conversation)	Differences by conversation
+ (1 | topic)	Differences by topics
+ (0 + interlocutor.baseline | speaker)	Per-speaker slope for convergence
+ (0 + interlocutor.baseline | interlocutor)	Per-interlocutor slope for convergence

with values that were highly correlated (Pearson $r > .97$) with the β_{S_b}. We therefore refitted the models without per-speaker intercepts. The full data for the six studies, as well as the models in which per-speaker intercepts were not omitted, are provided in the supplementary materials.

In every model, speakers were highly consistent in their production of that linguistic characteristic across conversations, as shown in Table 2, though consistency differed across the characteristics. Convergence, as indicated by the degree to which the interlocutor's baseline predicted for the variable, accounted for far less variance than the speaker's baseline, as Table 3 shows. However, significant convergence was found in every model. Fig. 3 illustrates speakers' consistency, and Fig. 4 illustrates their convergence, for each of the six studies.

Table 4 provides the standard deviation and model comparison-based p values for the three random intercepts that remained in the model. Note that these intercepts are identifying predictors of speakers' performance, not predictors of convergence, for example, whether F0 median was higher with particular interlocutors, conversations, or topics.

Conversation significantly contributed to predicting the speaker's performance in every model; there were indeed different patterns of each linguistic characteristic in each conversation. Because the conversation-level intercept applies to both conversation sides, it captures elements in the conversation in which the two participants vary together.

Table 2
Speaker consistency coefficients per characteristic

	β	*SE*	*df*	*t*	*p*
F0 median	0.971	0.004	3434	248.3	<.0001
F0 variance	0.675	0.012	3622	56.9	<.0001
Speech rate	0.800	0.009	4336	92.8	<.0001
uh:um ratio	0.787	0.009	4501	87.5	<.0001
Lexical information rate	0.645	0.009	4490	67.9	<.0001
Sentential conjunction	0.391	0.013	4677	29.9	<.0001

Table 3
Convergence coefficients per characteristic

	β	*SE*	*df*	*t*	*p*
F0 median	0.018	0.005	181	3.7	.0003
F0 variance	0.093	0.014	143	6.6	<.0001
Speech rate	0.048	0.010	219	4.9	<.0001
uh:um ratio	0.032	0.011	147	2.9	.0043
Lexical information rate	0.061	0.011	224	5.5	<.0001
Sentential conjunction	0.041	0.015	152	2.7	.0070

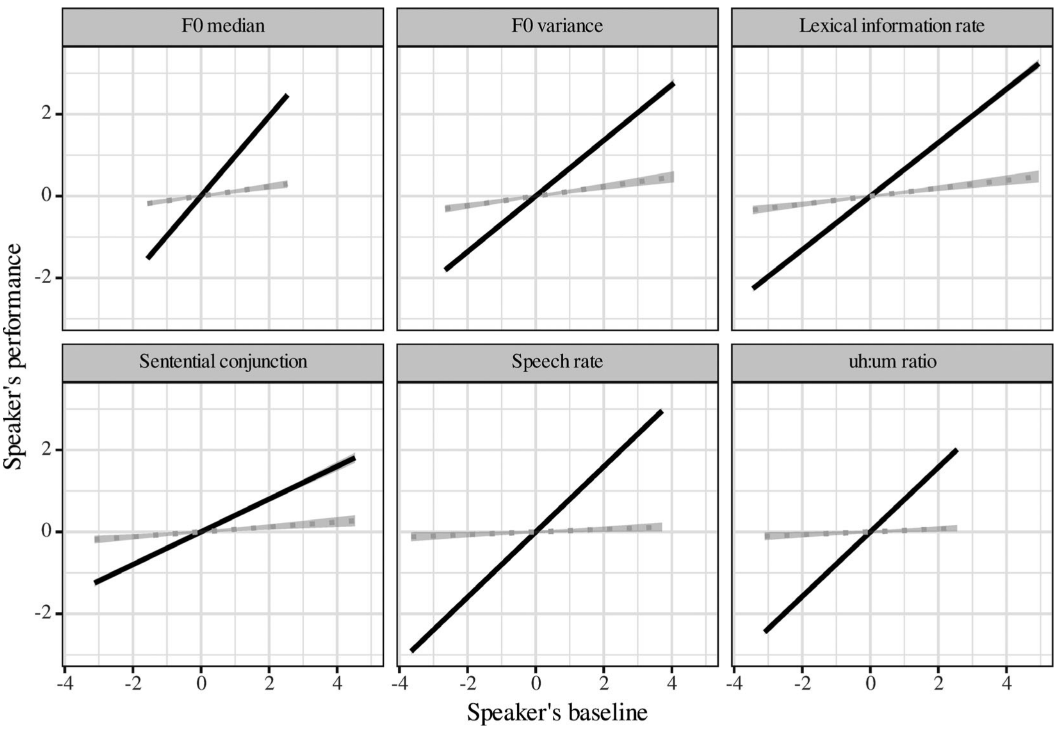

Fig. 3. By-characteristic consistency. The speaker's baseline (performance in other conversations), here on the *x*-axis, is strongly correlated with actual performance, here on the *y*-axis. The black line shows the linear trend. The standard error areas are not visible because they are so narrow that they do not extend beyond the width of the line. The dotted gray line shows the linear effect of the interlocutor's baseline, the measure of convergence, for comparison (with the interlocutor's baseline as the *x*-axis). A closer look at the effect of the interlocutor's baseline is given in Fig. 4, with an adjusted scale.

However, this is not necessarily convergence, as it may be driven by external characteristics, rather than speakers shifting toward each other. Similarly, topic contributed to every model, indicating that different topics have different typical values (e.g., higher pitch, more *uh* utterances).

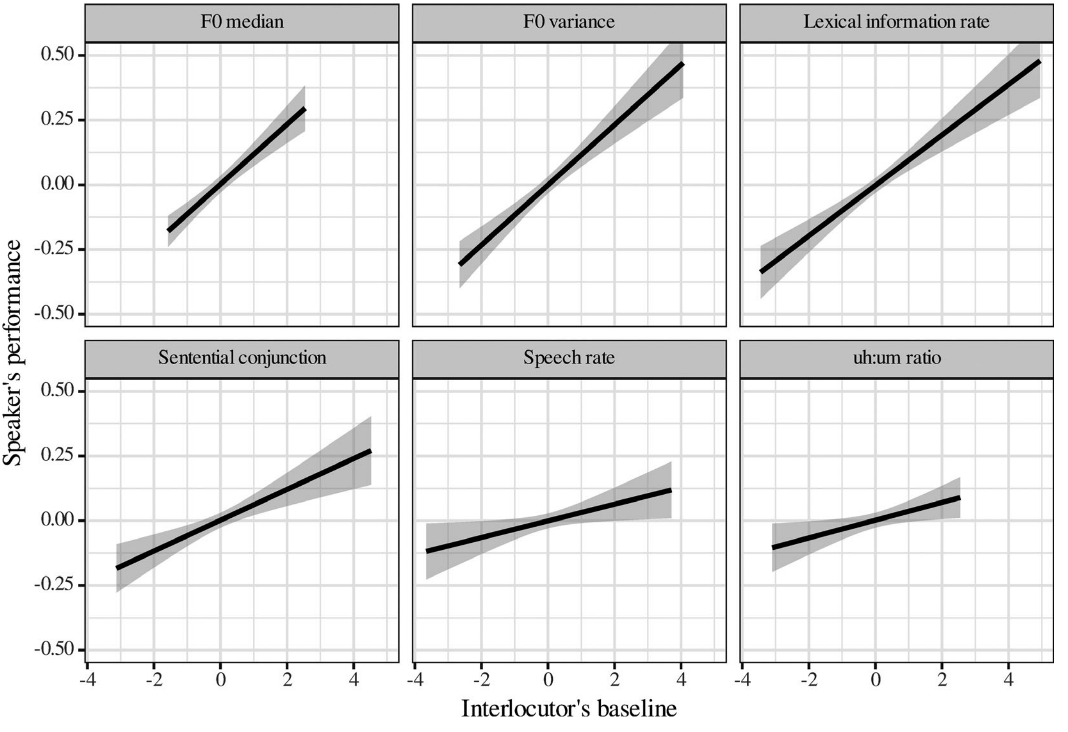

Fig. 4. By-characteristic convergence. The interlocutor's baseline (performance in other conversations), here on the *x*-axis, has a small but significant effect on speaker's performance, here on the *y*-axis. The line shows the linear trend, along with the standard error. We use a different scale for this plot than for the preceding plot because the convergence effects are much weaker than consistency.

Table 4
Random intercept *SD* and model comparison *p* values per characteristic

	Interlocutor		Topic		Conversation	
	SD	*p*	*SD*	*p*	*SD*	*p*
F0 median	0.0363	.0680	0.0218	.0283	0.0743	<.0001
F0 variance	0.0995	.0530	0.1227	<.0001	0.2602	<.0001
Speech rate	0.0907	.0012	0.1288	<.0001	0.2943	<.0001
uh:um ratio	0.0697	.1384	0.0509	.0240	0.2451	<.0001
Lexical information rate	0.0005	1.0000	0.4012	<.0001	0.2941	<.0001
Sentential conjunction	0.0697	.4568	0.2145	<.0001	0.2241	.0023

The interlocutor had a significant effect on speakers' performance only in the speech rate model, though marginal effects were found for several of the other characteristics. The lack of consistent per-interlocutor intercepts might suggest that the main influence that interlocutors have on speakers' performance is via convergence (measured with per-interlocutor slopes, and discussed below), rather than absolute effects.

Table 5
Random slope *SD* and model comparison *p* values per characteristic

	Per-Speaker Convergence Slope		Per-Interlocutor Conversation Slope	
	SD	*p*	*SD*	*p*
F0 median	0.0237	.2645	0.0298	.2153
F0 variance	0.0546	.4475	0.0608	.3209
Speech rate	0.0214	.8255	0.0000	1.0000
uh:um ratio	0.0734	.0173	0.0437	.4415
Lexical information rate	0.0000	.9998	0.0798	.0201
Sentential conjunction	0.0637	.3258	0.0629	.3829

Table 5 provides the standard deviation and model comparison-based *p* values for the per-speaker and per-interlocutor random slopes for the interlocutors' baselines in the six models. These values indicate whether there are individual tendencies in converging and eliciting convergence, respectively.

Per-speaker random slope was only significant for *uh*:*um* convergence, and per-interlocutor random slope was only significant for lexical information rate. This means that the individual-level variance in the convergence coefficients was largely negligible. These results suggest that individual differences in convergence by speaker and by interlocutor, if they exist, are too weak to be reliably detected by the method used here. Recent work has generally also found a lack of consistency in the degree of convergence exhibited by individual speakers across tasks (Pardo et al., 2018) or in different conversations (Cohen Priva & Sanker, 2018), though they exhibit some consistency within a linguistic characteristic in the same task (e.g., Wade et al., n.d.). Pardo et al. (2017) found that some model talkers consistently elicited more convergence than others in the same shadowing task with the same recordings. Consistency across conversations with the same individual (e.g., Sanker, 2015) could reflect either individual tendencies in producing convergence or tendencies in the degree of convergence elicited by a particular individual.

3. Cross-characteristic study

3.1. Introduction

While the by-characteristic analysis confirmed that convergence exists in all of the linguistic characteristics measured, there was only weak evidence for individual tendencies in producing convergence or eliciting convergence. These models also could not detect convergence patterns particular to each conversation, as each conversation has only two summary data points per characteristic (one for each speaker). In order to evaluate per-speaker, per-interlocutor, and per-conversation patterns in convergence as reflected across characteristics, we grouped together all six characteristics in a single model (they were already z-transformed, to facilitate comparison).

Previous studies have found replicable individual differences in various perceptual tasks (e.g., Ishida et al., 2016; Johnson, Watson, & Jensen, 1987; Kong & Edwards, 2016), and some work has found that personal traits such as AQ (Yu et al., 2013) and social desirability scores (Natale, 1975) are predictive of degree of convergence, but studies on individual differences in convergence have rarely used retesting to establish individual consistency in convergence. Individual tendencies in convergence may exist but be outweighed by larger effects such as the nature of each particular conversation or the speaker's opinion of the interlocutor. In this case, convergent patterns by conversation or by interlocutor may be more apparent than convergent tendencies of individuals. Such findings would support accounts of social factors mediating convergence (e.g., Babel, 2009; Giles et al., 1973; Pardo et al., 2012).

3.2. *Statistical models*

The six models in the by-characteristic analysis presented in the previous section each have a mixed effects model predicting the performance of the speakers based on their own baselines and the baselines of their interlocutors for a single linguistic characteristic. The cross-characteristic study presented here tests convergence across all characteristics. This provides up to 12 summary data points per conversation (six characteristics for each speaker); not all characteristics were available for all conversations, given exclusion procedures based on sound quality and the usage of each characteristic. As described above, each data point is a summary of the speaker's behavior in that characteristic across the entire conversation.

The regression analysis used each speaker's performance in a conversation as the dependent variable (once for each characteristic). The speaker's baseline in other conversations and the interlocutor's baseline in other conversations were used as the two main fixed predictors. The interlocutor's baseline in other conversations is the variable that measures convergence: A stronger effect of the interlocutor's baseline as a predictor of the speaker's productions indicates more convergence. These methods are the same as the ones that were used for the six individual characteristics in the by-characteristic analysis, except that each speaker in a conversation is repeated for every characteristic.

The main difference is in the structure of the random effects. The elements of the model are outlined below.

Several of the factors treat the linguistic characteristics as a pooled group; these factors are referred to as *characteristic-agnostic*, in contrast to the factors which test differences across characteristics, which are *per-characteristic*.

Several of the possible random intercepts would be meaningless in the context of the current study. For instance, it is meaningful to expect per-interlocutor intercepts to differ *within* each characteristic, but *characteristic-agnostic* per-interlocutor intercepts would arbitrarily group together different characteristics (e.g., higher F0 and increased use of *uh*). In contrast, characteristic-agnostic random slopes for convergence abstract over a putative single behavior, "participating in convergence," and should therefore be included in the model. As in the preliminary studies, we do not include random slopes for speaker

Table 6
The lme4 formula used to fit the model in the cross-characteristic study

lme4 Syntax	Explanation
Performance ~	Speaker's performance in a conversation
1	Global intercept (expected to be 0)
+ speaker.baseline	Speaker's baseline performance for the characteristic
+ interlocutor.baseline	Interlocutor's baseline performance for the characteristic
+ (0 + speaker.baseline | char)	Per-characteristic slope for consistency
+ (0 + interlocutor.baseline | char)	Per-characteristic slope for convergence
+ (0 + interlocutor.baseline | char:speaker)	Within-characteristic per-speaker slope for convergence
+ (0 + interlocutor.baseline | char:interlocutor)	Within-characteristic per-interlocutor slope for convergence
+ (1 | char:topic)	Within-characteristic per-topic differences
+ (1 | char:speaker)	Within-characteristic per-speaker differences
+ (1 | char:interlocutor)	Within-characteristic per-interlocutor differences
+ (1 | char:conversation)	Within-characteristic per-conversation differences
+ (0 + interlocutor.baseline | speaker)	Per-speaker slope for convergence (characteristic agnostic)
+ (0 + interlocutor.baseline | interlocutor)	Per-interlocutor slope for convergence (characteristic agnostic)
+ (0 + interlocutor.baseline | conversation)	Per-conversation slope for convergence (characteristic agnostic)

consistency, because this is not our variable of interest here. However, the supplementary materials provide information on speaker consistency slopes that do improve the model we present. The modeling choices are listed below, and they are summarized in Table 6.

Linguistic characteristic The model contains the per-characteristic random slopes, one for the speaker's baseline and one for the interlocutor's baseline. The two slopes are meant to account for the differences in consistency and convergence in different characteristics, as observed in the by-characteristic analysis. A per-characteristic intercept was not included, because the intercept was not significantly different than zero in the individual models for by-characteristic analysis.

Conversation We included a characteristic-agnostic random slope for convergence and a per-characteristic random intercept. We did not include a characteristic-agnostic random intercept, as explained above. Within-characteristic intercepts are meant to explain the variance that was captured by per-conversation intercepts in the by-characteristic analysis above. The random slope per conversation is one of our main variables of interest: If particular conversations elicit a higher or lower degree of convergence across multiple characteristics, this variable would significantly contribute to the model.

Speaker Despite its ultimate exclusion in the models for the individual characteristics, we included a within-characteristic random intercept per speaker. We also included a within-characteristic per speaker random slope for the interlocutors' baseline to model within-characteristic individual differences in convergence, as well as a characteristic-agnostic random slope for the interlocutors' baseline to model an overall tendency to converge. If there are individual tendencies in producing convergence, such that some speakers consistently converge more than other speakers ("convergers"), the slope would contribute to the model, making it better than a minimally different model that does not contain that slope. This slope is the second main variable of interest.

Interlocutor Interlocutor factors parallel the per-speaker random structure. We included a within-characteristic random intercept per-interlocutor, which mirrors the per-interlocutor random intercepts in the by-characteristic models. We also included both within-characteristic and characteristic-agnostic random slopes for convergence. The former mirrors the per-interlocutor random slope for convergence in the studies for individual characteristics, and the latter models the possibility that there are general tendencies in how much convergence particular interlocutors elicit, such that there is variance in the extent that interlocutors elicit convergence ("leadership"), across different speech characteristics. This is the third main variable of interest. As in the preliminary studies, being likely to elicit convergence and being likely to converge are distinct properties; a particular individual can have distinct likelihoods of each.

Topic For topic, we used a within-characteristic random intercept, which mirrors the per-topic intercept in the by-characteristic models.

3.3. Results and discussion

As expected, both the speakers' own baseline and their interlocutors' baseline performance were significant predictors of the speakers' performance ($\beta = 0.7109$, $SE = 0.079$, $df = 5.0$, $t = 8.95$, $p = .0003$; $\beta = 0.0471$, $SE = 0.011$, $df = 5.2$, $t = 4.45$, $p = .0060$, respectively): Speakers exhibited self-consistency in their productions across conversations, and they exhibited convergence. The results for the random effects are summarized in Table 7. All reported p values for random effects are adjusted for multiple comparisons (Benjamini & Hochberg, 1995).

Among the variables of interest, only the per-interlocutor slope for convergence significantly improved the model ($SD = 0.045$, $p = .0029$). This means that the model is improved when accounting for variance associated with differences in how much

Table 7
Summary of the cross-characteristic random effects. Raw p is the raw model comparison p value. FDR-adjusted p are the p values adjusted for multiple comparisons using false discovery rate (Benjamini & Hochberg, 1995)

	SD	Raw *p*	FDR-Adjusted *p*
Per-characteristic and conversation intercept	0.2474	<.0001	<.0001
Per-characteristic and interlocutor slope for convergence	0.0143	.8662	1.0000
Per-characteristic and interlocutor intercept	0.0715	.0005	.0015
Per-characteristic and speaker slope for convergence	0.0460	.0442	.0694
Per-characteristic and speaker intercept	0.0000	1.0000	1.0000
Conversation slope for convergence	0.0000	1.0000	1.0000
Interlocutor slope for convergence	0.0454	.0016	.0029
Speaker slope for convergence	0.0000	1.0000	1.0000
Per-characteristic and topic intercept	0.2011	<.0001	<.0001
Per-characteristic slope for convergence	0.0227	.0012	.0027
Per-characteristic slope for consistency	0.1943	<.0001	<.0001

convergence each interlocutor elicits, which provides evidence for varying degrees of leadership in convergence. We found no evidence for a per-speaker slope for convergence (SD = 0.000, p = 1.0000), suggesting that the effect of leadership in convergence is easier to find than an effect of individual propensity to converge. Similarly, we found no evidence for a per-conversation slope for convergence (SD = 0.000, p = 1.0000).

Fig. 5 shows the differences between interlocutors based on the per-interlocutor slope for convergence, grouped across linguistic characteristics and separated by linguistic characteristic. For readability, interlocutors are grouped into quartiles based on the degree of convergence they elicited. The first (lowest) quartile seems to show divergence: interlocutors from whom speakers tend to diverge. The other three quartiles show increasing degrees of convergence by quartile.

As expected, there were significant differences in per-characteristic slope for consistency and per-characteristic slope for convergence (SD = 0.194, p < .0001; SD = 0.023,

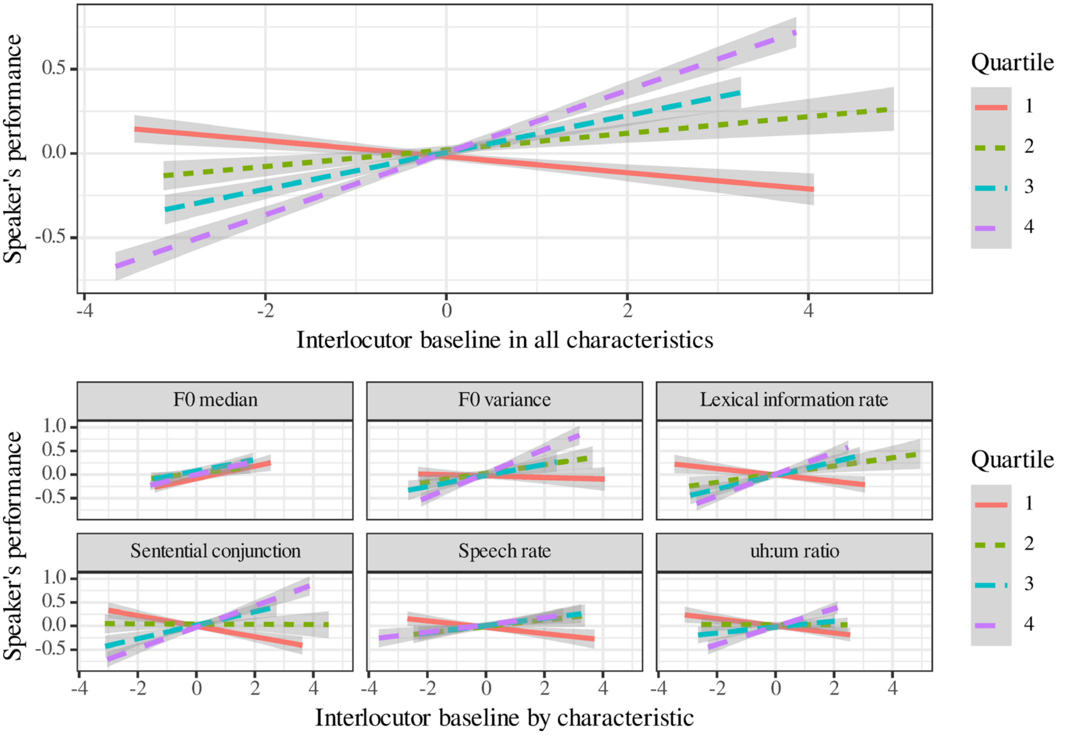

Fig. 5. By-quartile raw correlations between interlocutors' baseline and speakers' performance. Quartiles were determined based on the per-interlocutor slope for convergence. The top panel groups all the data points together, pooled across the six linguistic characteristics. The first quartile shows divergence, but the other three quartiles show increasing degrees of convergence. The bottom panel separates the correlations by characteristic. With the exception of F0 median, which does not show divergence in any quartile, all characteristics pattern much like the overall results, though quartiles 2–4 are less distinct from each other.

p = .0027, respectively); both self-consistency and convergence varied by linguistic characteristic. Similarly, as in the per-characteristic models, there were within-characteristic per-conversation differences and within-characteristic per-topic differences ($SD = 0.247$, $p < .0001$; $SD = 0.201$, $p < .0001$); there were absolute patterns in how each characteristic was realized based on the particular conversation and based on the conversation topic.

Unlike the per-characteristic models, the within-characteristic per-interlocutor differences did improve the model ($SD = 0.071$, $p = .0015$), as did the within-characteristic per-speaker slope for convergence (marginally, $SD = 0.046$, $p = .0694$). That is, there were per-speaker and per-interlocutor tendencies in convergence that were specific to the particular characteristic being measured. The preliminary studies only found weak per-speaker and per-interlocutor effects, using similar models. The reason for the difference could be differences in the variance captured by the fixed effects in the preliminary models and in this cross-characteristic model.

No other random effect had a significant effect on the model.

Although there was only a weak per-interlocutor effect in one of the separate models for convergence in each linguistic characteristic, there was a significant per-interlocutor effect in this cross-characteristic study. The combined study makes it possible to identify interlocutor effects that are definitely convergence (per-interlocutor slope for interlocutor performance as a predictor of the speakers' performance), rather than particular interlocutors motivating an absolute change in a particular characteristic (per-interlocutor intercept). In studies for individual speech characteristics, it can be difficult to distinguish between these factors, as both might produce the same effects. However, when comparing across speech characteristics, it will be clear if the shift elicited by the interlocutor reflects convergence or an absolute effect in a particular characteristic. For example, interlocutors who demonstrate difficulties understanding fast speech are likely to elicit slower speech, which might or might not align with the speech rate that those individuals use themselves; such effects would not be expected to predict convergence in other characteristics, because they are not actually measuring convergence. The influence of the interlocutor on degree of convergence across characteristics provides support for convergence being socially mediated at a high level.

Some of the per-interlocutor consistency in convergence can be explained by the social status or likeability of the interlocutor; speakers converge more to more socially desirable interlocutors (Gregory & Webster, 1996; Hwang & Chun, 2018) and to interlocutors whom they like more (Sanker, 2015; Yu et al., 2011). Our study does not aim to identify the particular traits that make some interlocutors elicit more convergence than others. However, Switchboard does provide demographic information about the speakers, so we did a post hoc test of whether age, gender, or college education predicted the per-interlocutor variation in convergence. None of these factors were significant, though there was a marginal tendency for older speakers to elicit less convergence. We also tested whether any of the six linguistic characteristics under investigation were predictors of per-interlocutor convergence (e.g., if speakers were more likely to converge to someone with more variable F0, slower speech rate, etc.); none of these factors were significant (for the model testing all of these factors, see the Supplementary Materials, Section 3). This model also tested whether amount of speech produced by the interlocutor, either counted by word or by time, predicted per-interlocutor convergence; neither factor was a significant predictor of how much convergence each interlocutor elicited, indicating that per-interlocutor variation in convergence is not due to variation in talkativeness and consequently amount of exposure.

In contrast to the per-interlocutor slope, the per-speaker convergence slope did not improve the model. Given that the method was sensitive enough to find support for per-interlocutor tendencies in convergence, this null result is not likely to be due to insufficient power. Individual propensity to converge was not strong enough for our model to detect, providing little support for the view of variation in convergence as reflecting inherent individual traits. The results instead support the argument that variation in degree of convergence largely reflects properties of the interaction rather than aspects of the individual. Notably, per-interlocutor convergence is not a form of individual differences in the sense of individual tendencies in production or perception. Interlocutors who elicit greater convergence have traits that make others converge *to them*; this effect does not demonstrate that these individuals have any tendencies in producing convergence themselves.

4. General discussion

Our paper addresses two main questions: Is there evidence for individual tendencies in convergence, reflected similarly across different linguistic characteristics, which would indicate a broad cognitive trait responsible for convergence? Is there evidence for consistent effects of particular interlocutors in eliciting convergence, which would indicate social effects in mediating convergence? Both of these issues fit into a larger question of how convergence arises and what drives variation in convergence.

A key aspect of our study is examining convergence across several linguistic characteristics rather than in a single characteristic. The degree of convergence found even within the same task or conversation can depend on the characteristic used to measure it (Babel, 2012; Levitan & Hirschberg, 2011; Pardo et al., 2017), which could suggest that there are functional differences in convergence for different characteristics, or in how speakers represent them. However, simpler explanations have also been proposed. If differences in convergence reflect differences in exposure, speakers should converge more to characteristics which they have heard more examples of (Kaschak et al., 2011; Oben & Brône, 2016); our results do not show clear evidence for greater convergence in characteristics with ongoing exposure (like F0) than in characteristics with fewer tokens (like the *uh:um* ratio). If differences in convergence reflect differences in variability, with more convergence to more variable characteristics (Babel, 2009; Pierrehumbert, 2002), then there should be more convergence in characteristics for which individual consistency is lower; our results do not exhibit such a correlation. While differences in salience of each characteristic (Johnson, 1997; Pardo et al., 2012) provide a possible source of variation in convergence, this factor cannot be clearly evaluated given the lack of measurement for relative salience for each of the linguistic characteristics tested in this study. Despite absolute differences in convergence, some of the same effects are apparent across different characteristics. The per-interlocutor effects across characteristics indicate that social mediation influences convergence in different characteristics in the same ways. The lack of evidence for individual tendencies in convergence leaves open the possibility that there are some underlying differences in convergence in different characteristics.

4.1. Speaker effects in convergence

Within our data, we found no evidence for a contribution of individual tendencies in predicting convergence across different linguistic characteristics. The interlocutor was a significant predictor of degree of convergence in our data, with exactly the same sample size as the speaker, as the same individuals were considered in both roles; thus, the lack of evidence for individual tendencies in convergence is unlikely to result from inadequate sample size.

Individual differences in perception have been demonstrated previously, and they can have high test–retest reliability with the same or closely related tests, while correlations across less closely related tests tend to be weak and the differences do not correlate well with performance in most cognitive tests (e.g., Kidd, Watson, & Gygi, 2007; Surprenant & Watson, 2001). In perception and phonological processing, individual differences have been identified for characteristics such as categoricalness of phonological perception (Kapnoula, 2016; Kong & Edwards, 2016), compensation for coarticulation with the phonological environment (Repp, 1981; Yu, 2010; Yu & Lee, 2014), accuracy of perception of noise (Lev-Ari, 2018b), weighting of phonetic cues for particular phonological contrasts (Schertz et al., 2015; Schultz et al., 2012), and degree of influence of the lexicon on phonological perception (Stewart & Ota, 2008).

There is more limited evidence for individual consistency in convergence. Some traits of the speaker are predictive of degree of convergence, for example, AQ (Yu et al., 2013), social desirability scores (Natale, 1975), social network size (Lev-Ari, 2018a), and tendency to compromise (Weatherholtz et al., 2014). However, few convergence studies include retesting to demonstrate that individuals are consistent in these differences, and the size of the effect is often small. There is some evidence for individual tendencies in convergence within a given linguistic characteristic between instances of same task or similar tasks (Sanker, 2015; Wade et al., n.d.), but evidence for tendencies across different tasks is weaker (Pardo et al., 2018). The results of our study, in which each individual conversed with several different interlocutors on different topics, provide little evidence for individual tendencies in convergence. If there are individual differences in convergence, as has been suggested by previous studies, these differences should be apparent when speakers participate in several conversations. The per-interlocutor tendencies in convergence, which are discussed in the following section, demonstrate that the effect of the interlocutor is similar across different partners, so any variation due to the speaker could be identified separately. If speakers exhibit no tendencies in convergence that hold across different tasks or conversations with different partners, this substantially constrains the interpretations that can be given to apparent variation.

While some individual tendencies in convergence may exist, our results suggest that they do not reflect any broad cognitive trait (cf. Yu, 2013; Yu et al., 2013), but rather must be specific to particular linguistic characteristics. Little previous work has tested correlations across characteristics; the work that has looked for them has generally found no correlations (Bilous & Krauss, 1988; Pardo et al., 2012; Sanker, 2015; Weise & Levitan, 2018), with the occasional correlation that can probably be attributed to physical relationships between characteristics (Cohen Priva & Sanker, 2018) or repeated measures producing a false positive (Rahimi, Kumar, Litman, Paletz, & Yu, 2017; Sanker, 2015). In studies with a relatively small number of participants, a null result for individual tendencies can be difficult to interpret; there has been no previous large-scale comparison across characteristics, controlling for effects of particular conversations. Our study, using a corpus with a large number of speakers, provides clearer evidence for a lack of individual convergent tendencies that hold across linguistic characteristics. The absence of a significant effect in this large corpus indicates that if broad individual differences in convergence do exist, they would be too weak to be detected in normal experimental settings.

4.2. Interlocutor effects in convergence

The results of our study indicate that variation in convergence is influenced by the interlocutor: Individual interlocutors have tendencies in how much convergence they elicit. Previous studies with smaller sets of individuals have similarly demonstrated variation based on the interlocutor or model talker (e.g., Hwang & Chun, 2018; Pardo et al., 2017), and effects of particular manipulations of the actual or perceived traits of the interlocutor or model talker, such as interlocutor status (e.g., Bane et al., 2010; Gregory & Webster, 1996) and the speaker's opinion of the interlocutor or model talker (e.g., Bourhis & Giles, 1977; Pardo et al., 2012; Yu et al., 2011). We find that interlocutor-specific variation in convergence is consistent across linguistic characteristics; that is, the factors that strengthen or weaken convergence to a particular interlocutor in one linguistic characteristic similarly affect other characteristics.

It is noteworthy that the variation in how much convergence individuals elicited did not just range from no convergence to substantial convergence, but also included individuals who elicited divergence. This pattern was apparent in each of the linguistic characteristics individually, except for F0 median. Most studies find that convergence occurs even in socially neutral contexts like repeating after recordings of isolated words in a lab and similarly find that even with interlocutors or conditions that elicit less convergence, the low ranges are near zero convergence, rather than divergence (e.g., Babel, 2009; Pardo et al., 2017). Convergence might be the default automatic response to linguistic input, with inhibition of convergence requiring more conscious effort (Street & Giles, 1982). Speakers can deliberately modify such characteristics as a way of asserting identity, particularly with contrastive dialect features (Bourhis & Giles, 1977), but they may still converge to some degree even when they have negative opinions about speakers of the dialect they are hearing (e.g., Babel, 2010). If convergence needs to be actively suppressed, divergence should only be possible for linguistic characteristics which are under more conscious control rather than being more automatic. If subsequent work investigates additional sub-phonological phonetic characteristics, it is likely that such characteristics will behave like F0 median does in our results, exhibiting little or no divergence.

The presence of strong interlocutor effects within our data suggests that convergence is socially mediated, even if the mechanism driving convergence is not inherently social. Speakers converge more to speakers whom they like based either on the context of their interaction (Bourhis & Giles, 1977; Yu et al., 2011) or whom they perceive as high status (Bane et al., 2010; Gregory & Webster, 1996), and they are also more likely to converge to voices that are perceived as more standard (Weatherholtz et al., 2014) or more attractive (Babel et al., 2014). Per-interlocutor consistency in convergence, even across different linguistic characteristics and across conversations with different partners and different conversational topics, suggests that the social effects occur at a high level, perhaps reflecting interspeaker alignment (cf. Jiang et al., 2015; Pickering & Garrod, 2004), with different interlocutors eliciting different degrees of alignment. The variation in degree of convergence with different interlocutors could result from differences in the weight given to input from the particular interlocutor; Chartrand and Bargh (1999) make a similar proposal, with more convergence resulting from greater perceptual engagement with the interlocutor. Because convergence provides a small-scale illustration of the propagation of change (Yu, 2013), our finding of per-interlocutor variation in convergence helps inform how social variation might structure which innovations spread, in addition to contributing to our understanding of social dimensions of convergence.

Variation in convergence elicited by each interlocutor cannot be attributed to how much exposure to the interlocutor speakers received. Within our data, per-interlocutor convergence was not correlated with the average amount of speech produced by that interlocutor, either as measured in number of words or total time spent talking. Consistent with this lack of correlation, variation in convergence elicited by different interlocutors has also been observed in previous studies which provided equal exposure to each voice (e.g., Hwang & Chun, 2018; Pardo et al., 2017).

4.3. Conversation effects in convergence

In contrast with interlocutor effects, the particular conversation was not a significant predictor of convergence. Some previous work has found that degree of convergence is predicted by aspects of the particular relationship between interlocutors (e.g., Bane et al., 2010; Pardo et al., 2012; Sanker, 2015). However, such work has often looked at each participant in a single interaction, so effects of the interlocutors and effects of the particular conversation could not be separated out, except indirectly based on finding differences associated with traits of the speakers, aspects of the conversation, or experimental conditions manipulating aspects of the conversation. While our results do not provide evidence for effects of the conversation, we have less data for each conversation than for each speaker or interlocutor, so the lack of per-conversation effect in our results might simply be the result of insufficient data to capture such an effect.

5. Conclusion

We present new data on convergence across linguistic characteristics and across speakers, using a large corpus of natural speech. Although there is variation across characteristics in convergence and in speakers' consistency, different characteristics exhibited shared influences of the interlocutor, which provides evidence for high-level mediation of convergence, reflected across different linguistic characteristics.

There was little evidence for by-speaker tendencies in how much each individual converged; to the extent that individual tendencies in convergence exist, they are specific to particular linguistic characteristics, rather than reflecting a broad cognitive trait. Thus, individual tendencies in convergence are unlikely to provide strong insights into individual linguistic variation in broad traits, but they might inform variation in behavior of specific details.

The interlocutor was predictive of degree of convergence across linguistic characteristics and across conversations, indicating that some individuals are more likely to elicit convergence than others. This per-interlocutor variation in eliciting convergence demonstrates the importance of situational factors in mediating convergence. Moreover, it suggests that this mediation occurs at a high level.

Acknowledgments

We thank Elinor Amit, Molly Babel, Roman Feiman, Jim Morgan, Jennifer Pardo, Antje Schweitzer, Meredith Tamminga, and the audience of the Workshop on Accommodation in Speech Communication for helpful discussions and feedback on earlier versions of this paper, as well as the three anonymous reviewers who helped us refine the paper.

Open Research badges

This article has earned Open Data and Open Materials badges. Data and materials are available at https://osf.io/bdx9q/.

Notes

1. Convergence, as thus defined, only includes alignment between individuals, and not other sorts of interlocutor accommodation. Convergence can be measured with two basic types of comparisons: overall increased similarity of a speaker to an interlocutor or model talker (e.g., Babel, 2009; Goldinger, 1998; Pardo et al., 2017) and synchrony over time (e.g., Levitan & Hirschberg, 2011; Schweitzer & Lewandowski, 2013). Our study only examines the former.
2. The cross-characteristics study adds a per-conversation random slope for β_R, which cannot be included in the preliminary studies because there are too few observations per conversation.
3. We did run such models, and though the models took longer to converge and were more redundant than the models we present here, the results were essentially the same as the models we present here. The sentential conjunction model (described below) did not converge. The other models can be found in the supplementary materials.
4. It is impossible to use a fully maximal random effects structure for every fixed effect in most models, as has been shown by Bates, Kliegl, Vasishth, and Baayen (2018) and Matuschek et al. (2017). Adding as many random effects as possible increases Type 2 errors and is likely to produce a model that does not converge; indeed, including all possible random effects in our data results in a model that does not converge. It then becomes necessary to come up with a system for choosing which random effects will be kept and which will be omitted, usually in a way which reflects the design of the study and the question being investigated. In the supplementary materials, we present models with more extensive random effects, along with a discussion of alternative ways that the random effects structure could be selected.
5. Excluding the most frequent words does not have a substantial effect on the model results.

References

Abel, J., & Babel, M. (2017). Cognitive load reduces perceived linguistic convergence between dyads. *Language and Speech*, *60*(3), 479–502. https://doi.org/10.1177/0023830916665652

Acton, E. K. (2011). On gender differences in the distribution of Um and Uh. *University of Pennsylvania Working Papers in Linguistics*, *17*(2), (2).

Babel, M. (2009). *Phonetic and social selectivity in speech accommodation*. PhD thesis, University of California, Berkeley.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, *39*, 437–456. https://doi.org/10.1017/S0047404510000400

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*(1), 177–189. https://doi.org/10.1016/j.wocn.2011.09.001

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

Babel, M., McGuire, G., Walters, S., & Nicholls, A. (2014). Novelty and social preference in phonetic accommodation. *Laboratory Phonology*, *5*(1), 123–150. https://doi.org/10.1515/lp-2014-0006

Baker, A., Archangeli, D., & Mielke, J. (2011). Variability in American English s-retraction suggests a solution to the actuation problem. *Language Variation and Change*, *23*(3), 347–374. https://doi.org/10.1017/S0954394511000135

Bane, M., Graff, P., & Sonderegger, M. (2010). Longitudinal phonetic variation in a closed system. *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, *46*, 43–58.

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. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., & Baayen, R. H. (2018). *Parsimonious mixed models*. arXiv:1506.04967v2 [stat.ME]. https://arxiv.org/abs/1506.04967v2

Beddor, P. S., Coetzee, A. W., Styler, W., McGowan, K. B., & Boland, J. E. (2018). The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. *Language*, *94*(4), 931–968. https://doi.org/10.1353/lan.2018.0051

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

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. https://doi.org/10.1016/0271-5309(88)90016-X

Bourhis, R. Y., & Giles, H. (1977). The language of intergroup distinctiveness. In H. Giles (Ed.), *Language, ethnicity and intergroup relations* (pp. 119–136). New York: Academic Press.

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. https://doi.org/10.1037/0022-3514.76.6.893.

Cieri, C., Graff, D., Kimball, O., Miller, D., & Walker, K. (2005). *Fisher English training part 2, transcripts*. Philadelphia, PA: Linguistic Data Consortium.

Cieri, C., Miller, D., & Walker, K. (2004). The Fisher corpus: A resource for the next generations of speech-to-text. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC)* (pp. 69–71). Lisbon, Portugal: European Language Resources Association (ELRA).

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

Cohen Priva, U. (2017). Not so fast: Fast speech correlates with lower lexical and structural information. *Cognition*, *160*, 27–34. https://doi.org/10.1016/j.cognition.2016.12.002

Cohen Priva, U., Edelist, L., & Gleason, E. (2017). Converging to the baseline: Corpus evidence for convergence in speech rate to interlocutor's baseline. *Journal of the Acoustical Society of America*, *141*(5), 2989–2996. https://doi.org/10.1121/1.4982199

Cohen Priva, U., & Jaeger, T. F. (2018). The interdependence of frequency, predictability, and informativity. *Linguistics Vanguard*, *4*. https://doi.org/10.1515/lingvan-2017-0028

Cohen Priva, U., & Sanker, C. (2018). Distinct behaviors in convergence across measures. In C. Kalish M. Rau J. Zhu & T. Rogers (Eds.), *Proceedings of the 40th Annual Meeting of the Cognitive Science Society* (pp. 1515–1520). Madison, WI: Cognitive Science Society.

Cohen Priva, U., & Sanker, C. (2019). Limitations of difference-in-difference for measuring convergence. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *10*(1), 15. https://doi.org/10.5334/labphon.200

Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica*, *64*(2–3), 145–173. https://doi.org/10.1159/000107914

Dingemanse, N. J., Kazem, A. J. N., Réale, D., & Wright, J. (2010). Behavioural reaction norms: Animal personality meets individual plasticity. *Trends in Ecology & Evolution*, *25*(2), 81–89. https://doi.org/10.1016/j.tree.2009.07.013

Dorgeloh, H. (2004). Conjunction in sentence and discourse: Sentence-initial *and* and discourse structure. *Journal of Pragmatics*, *36*(10), 1761–1779. https://doi.org/10.1016/j.pragma.2004.04.004

Drager, K., & Hay, J. (2012). Exploiting random intercepts: Two case studies in sociophonetics. *Language Variation and Change*, *24*(1), 59–78. https://doi.org/10.1017/S0954394512000014

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. https://doi.org/10.1016/S0749-596X(03)00072-X

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. https://doi.org/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 (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1–68). Cambridge, UK: Cambridge University Press.

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

Godfrey, J. J., & Holliman, E. (1997). *Switchboard-1 release 2*. Philadelphia: Linguistic Data Consortium.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279. https://doi.org/10.1037/0033-295X.105.2.251

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. https://doi.org/10.1037/0022-3514.70.6.1231

Grosvald, M., & Corina, D. (2012). The production and perception of sub-phonemic vowel contrasts and the role of the listener in sound change. In M.-J. Solé & D. Recasens (Eds.), *The initiation of sound change: Perception, production, and social factors* (pp. 77–100). Amserdam; Philadelphia: John Benjamins.

Harkins, D., Feinstein, D., Lindsey, T., Martin, S., & Winter, G. (2003). *Switchboard MS State manually corrected word alignments*. Available at: https://www.isip.piconepress.com/projects/switchboard/. Accessed Janurary 2016.

Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *Journal of the Acoustical Society of America*, *123*(5), 2825–2835. https://doi.org/10.1121/1.2897042

Heath, J. (2017). How automatic is phonetic convergence? Evidence from working memory. *Proceedings of the Linguistic Society of America*, *2*, Art. 35. https://doi.org/10.3765/plsa.v2i0.4088

Hornickel, J., Skoe, E., Nicol, T., Zecker, S., & Kraus, N. (2009). Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(31), 13022–13027. https://doi.org/10.1073/pnas.0901123106

Hwang, H., & Chun, E. (2018). Influence of social perception and social monitoring on structural priming. *Cognitive Science*, *42*, 303–313. https://doi.org/10.1111/cogs.12604

Ishida, M., Samuel, A. G., & Arai, T. (2016). Some people are "more lexical" than others. *Cognition*, *151*, 68–75. https://doi.org/10.1016/j.cognition.2016.03.008

Jiang, J., Chen, C., Dai, B., Shi, G., Ding, G., Liu, L., & Lu, C. (2015). Leader emergence through interpersonal neural synchronization. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(14), 4274–4279. https://doi.org/10.1073/pnas.1422930112

Johnson, D. M., Watson, C. S., & Jensen, J. K. (1987). Individual differences in auditory capabilities. I. *Journal of the Acoustical Society of America*, *81*(2), 427–438. https://doi.org/10.1121/1.394907

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–165). New York: Academic Press.

Jun, S.-A., & Bishop, J. (2015). Prominence in relative clause attachment: Evidence from prosodic priming. In L. Frazier & E. Gibson (Eds.), *Explicit and implicit prosody in sentence processing: Studies in honor of Janet Dean Fodor* (pp. 217–240). New York: Springer.

Kapnoula, E. E. (2016). *Individual differences in speech perception: Sources, functions, and consequences of phoneme categorization gradiency*. PhD thesis, University of Iowa.

Kaschak, M. P., Kutta, T. J., & Jones, J. L. (2011). Structural priming as implicit learning: Cumulative priming effects and individual differences. *Psychonomic Bulletin & Review*, *18*(6), 1133–1139. https://doi.org/10.3758/s13423-011-0157-y

Kataoka, R. (2011). *Phonetic and cognitive bases of sound change*. PhD thesis, University of California, Berkeley.

Kidd, G. R., Watson, C. S., & Gygi, B. (2007). Individual differences in auditory abilities. *Journal of the Acoustical Society of America*, *122*(1), 418–435. https://doi.org/10.1121/1.2743154

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. https://doi.org/10.1515/labphon.2011.004

Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics*, *59*, 40–57. https://doi.org/10.1016/j.wocn.2016.08.006

Labov, W. (2001). *Principles of linguistic change, vol. 2: Social factors*. Oxford: Blackwell.

Large, N. R., Frisch, S., & Pisoni, D. B. (1998). Perception of wordlikeness: Effects of segment probability and length on subjective ratings and processing of non-word sound patterns. *Research on Spoken Language Processing: Progress Report*, *22*, 95–125.

Lev-Ari, S. (2018a). Social network size can influence linguistic malleability and the propagation of linguistic change. *Cognition*, *176*, 31–39. https://doi.org/10.1016/j.cognition.2018.03.003

Lev-Ari, S. (2018b). The influence of social network size on speech perception. *Quarterly Journal of Experimental Psychology*, *71*(10), 2249–2260. https://doi.org/10.1177/1747021817739865

Levitan, R., & Hirschberg, J. B. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *Proceedings of Interspeech*, *5*, 3081–3084.

Martin, J. G. A., Nussey, D. H., Wilson, A. J., & Réale, D. (2011). Measuring individual differences in reaction norms in field and experimental studies: A power analysis of random regression models. *Methods in Ecology and Evolution*, *2*(4), 362–374. https://doi.org/10.1111/j.2041-210X.2010.00084.x

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

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. https://doi.org/10.1016/j.cognition.2008.08.002

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. https://doi.org/10.1177/026192702237958

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, *32*(5), 790–804. https://doi.org/10.1037/0022-3514.32.5.790

Oben, B., & Brône, G. (2016). Explaining interactive alignment: A multimodal and multifactorial account. *Journal of Pragmatics*, *104*, 32–51. https://doi.org/10.1016/j.pragma.2016.07.002

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

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

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

Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics*, *79*, 637–659. https://doi.org/10.3758/s13414-016-1226-0

Pardo, J. S., Urmanche, A., Wilman, S., Wiener, J., Mason, N., Francis, K., & Ward, M. (2018). A comparison of phonetic convergence in conversational interaction and speech shadowing. *Journal of Phonetics*, *69*, 1–11. https://doi.org/10.1016/j.wocn.2018.04.001

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., & Zandipour, M. (2004). The distinctness of speakers? Productions of vowel contrasts is related to their discrimination of the contrasts. *Journal of the Acoustical Society of America*, *116*(4), 2338–2344. https://doi.org/10.1121/1.1787524

Pickering, M. J., & Garrod, S. (2004). Towards a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(2), 169–190. https://doi.org/10.1017/S0140525X04000056

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*, 329–392. https://doi.org/10.1017/S0140525X12001495

Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology VII* (pp. 101–140). Berlin: Mouton de Gruyter.

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). *Buckeye corpus of conversational speech*. Columbus: Department of Psychology, Ohio State University.

Rahimi, Z., Kumar, A., Litman, D., Paletz, S., & Yu, M. (2017). Entrainment in multi-party spoken dialogues at multiple linguistic levels. *Proceedings of Interspeech*, *3*, 1696–1700.

Reitter, D., Moore, J. D., & Keller, F. (2006). Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *28*, 685–690.

Repp, B. H. (1981). Two strategies in fricative discrimination. *Perception & Psychophysics*, *30*(3), 217–227. https://doi.org/10.3758/BF03214276

Sanker, C. (2015). Comparison of phonetic convergence in multiple measures. In *Cornell working papers in phonetics and phonology 2015* (pp. 60–75). https://doi.org/10.5281/zenodo.3726190

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

Schertz, J., & Clare, E. J. (2020). Phonetic cue weighting in perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(2), e1521. https://doi.org/10.1002/wcs.1521

Schiffrin, D. (1986). Functions of *and* in discourse. *Journal of Pragmatics*, *10*, 41–66. https://doi.org/10.1016/0378-2166(86)90099-8

Schultz, A. A., Francis, A. L., & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *Journal of the Acoustical Society of America*, *132*(2), EL95–EL101. https://doi.org/10.1016/j.wocn.2017.09.009

Schweitzer, A., & Lewandowski, N. (2013). Convergence of articulation rate in spontaneous speech. *Proceedings of Interspeech*, *1*, 525–529.

Song, J. H., Nicol, T., & Kraus, N. (2011). Test–retest reliability of the speech-evoked auditory brainstem response. *Clinical Neurophysiology*, *122*(2), 346–355. https://doi.org/10.1016/j.clinph.2010.07.009

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). A scale for the measurement of psychological magnitude pitch. *Journal of the Acoustical Society of America*, *8*, 185–190. https://doi.org/10.1121/1.1915893

Stewart, M. E., & Ota, M. (2008). Lexical effects on speech perception in individuals with “autistic” traits. *Cognition*, *109*(1), 157–162. https://doi.org/10.1016/j.cognition.2008.07.010

Street, R. L. J., & Giles, H. (1982). Speech accommodation theory: A social cognitive approach to language and speech behavior. In M. E. Roloff & C. R. Berger (Eds.), *Social cognition and communication* (pp. 193–226). Beverly Hills: Sage.

Surprenant, A. M., & Watson, C. S. (2001). Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners. *Journal of the Acoustical Society of America*, *110*(4), 2085–2095. https://doi.org/10.1121/1.1404973

Wade, L., Lai, W., & Tamminga, M. (n.d.). The reliability of individual differences in VOT imitation. *Language and Speech*.

Weatherholtz, K., Campbell-Kibler, K., & Jaeger, T. F. (2014). Socially mediated syntactic alignment. *Language Variation and Change*, *26*, 387–420. https://doi.org/10.1017/S0954394514000155

Weise, A., & Levitan, R. (2018). Looking for structure in lexical and acoustic-prosodic entrainment behaviors. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, *2*, 297–302.

Yu, A. C. L. (2010). Perceptual compensation is correlated with individuals' "autistic" traits: Implications for models of sound change. *PLoS ONE*, *5*(8), e11950. https://doi.org/10.1371/journal.pone.0011950

Yu, A. C. L. (2013). Individual differences in socio-cognitive processing and the actuation of sound change. In A. C. L. Yu (Ed.), *Origins of sound change: Approaches to phonologization* (pp. 201–227). Oxford, UK: Oxford University Press.

Yu, A. C. L. (2019). On the nature of the perception-production link: Individual variability in English sibilant-vowel coarticulation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *10*(1), 2. https://doi.org/10.5334/labphon.97

Yu, A. C. L., Abrego-Collier, C., Baglini, R., Grano, T., Martinovic, M., Otte III, C., Thomas, J., & Urban, J. (2011). Speaker attitude and sexual orientation affect phonetic imitation. *University of Pennsylvania Working Papers in Linguistics*, *17*(1), 235–242.

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. https://doi.org/10.1371/journal.pone.0074746

Yu, A. C. L., & Lee, H. (2014). The stability of perceptual compensation for coarticulation within and across individuals: A cross-validation study. *Journal of the Acoustical Society of America*, *136*(1), 382–388. https://doi.org/10.1121/1.4883380

Yu, A. C. L., & Zellou, G. (2019). Individual differences in language processing: Phonology. *Annual Review of Linguistics*, *5*, 131–150. https://doi.org/10.1146/annurev-linguistics-011516-033815

Yun, K., Watanabe, K., & Shimojo, S. (2012). Interpersonal body and neural synchronization as a marker of implicit social interaction. *Scientific Reports*, *2*, Art. 959. https://doi.org/10.1038/srep00959

Zellou, G. (2017). Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics*, *61*, 13–29. https://doi.org/10.1016/j.wocn.2016.12.002

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Appendix S1. Supplementary materials.