Speech Communication 131 (2021) 23–34

Contents lists available at ScienceDirect

Speech Communication

journal homepage: www.elsevier.com/locate/specom

Who converges? Variation reveals individual speaker adaptability

Yoonjeong Lee [a,*], Louis Goldstein [b], Benjamin Parrell [c], Dani Byrd [b]

[a] University of California, Los Angeles, Los Angeles, 90095 CA, United States
[b] University of Southern California, Los Angeles, 90007 California, United States
[c] University of Wisconsin-Madison, Madison, 53706 Wisconsin, United States

ARTICLE INFO

Keywords:
Speech accommodation
Variability
Adaptability
Convergence
Flexibility
Conversational interaction

ABSTRACT

Little is known about the cognitive capacities underlying real-time accommodation in spoken language and how they may allow conversing speakers to adapt their speech production behaviors. This study first presents a simple attunement model that incorporates hypothesized capacities, with a focus on individual variability as one of those capacities. The model makes explicit predictions about observable convergence behaviors in interacting speakers, including that: i) the intrinsically more variable speaker of the two will be the one who converges to their partner, ii) this flexible speaker with higher baseline variability will exhibit a substantial decrease in variability and iii) a greater change in the variability between speaking solo and interacting with their partner. These predictions are supported by the results of the modeling simulations. To further test the model's predictions, we analyzed a behavioral dataset including acoustic and articulatory data from three pairs of interacting speakers participating in a maze navigation task as well as a like solo speech task. The amount of variability in the speech parameters of each dyad member was quantified using coefficient of variation. The experimental results parallel the simulation results, and taken together, this work indicates that structured variability is an illuminating index of individual speaker adaptability and convergence behavior.

1. Introduction

A hallmark property of healthy speech production is its adaptability. Speakers can quickly and effortlessly adapt the movements and orchestration of their articulators as a function of the task demands of the speaking situation, such as those adjustments involved in whispering versus yelling, speaking casually to friends versus giving a lecture, reading poetry versus singing, or giving instructions versus answering questions (Gordon Danner, Vilela Barbosa and Goldstein 2018). Less obviously perhaps, recent research has shown that speakers adapt, or accommodate, their speech to that of their interlocutor, resulting in a tendency towards convergence between the two speakers' production patterns (Abney et al., 2014, Kello and Warlaumont 2015; Babel 2012; Cohen Priva, Edelist and Gleason 2017; Giles 1973, 2008; Goldinger 1998; Levitan and Hirschberg 2011; Nielsen 2011; Pardo 2006, among others), or even sometimes towards divergence (Bourhis and Giles 1977; Lee et al., 2018; Pardo et al., 2012). Although much of the evidence for accommodation has come from acoustic measures of speech production, some studies have also shown that accommodation can occur in speakers' articulatory movements (Lee et al., 2018; Tiede et al., 2010; Tiede and Mooshammer 2013; Vatikiotis-Bateson, Barbosa and Best, 2014) and the dynamical control parameters underlying those movements (Lee et al., 2018).

While there is ample empirical evidence for accommodation in speech production, how the cognitive system implements this adaptability remains an outstanding question in understanding spoken language interaction. Accommodation phenomenon between speakers is particularly challenging to understand because the adaptations involved require dynamic, real-time changes in control of the speech articulators, unlike the adaptations to other sorts of task demands, like giving a public lecture, that can involve deployment of well-learned styles of speaking. What cognitive capacities allow speakers to make these real-time adaptations?

One key component of this capacity lies in speakers' *variability in the production of phonetic units*. Variability in the values of the parameters that identify the goal/target for a particular phonological unit has been observed both within and across speakers (e.g., Johnson et al., 1993; Whalen et al., 2018). For example, Harper (2020) found substantial within-speaker variability even when examining a specific phonetic unit produced in the same phonetic context and at the same nominal

* Corresponding author.
E-mail address: yoonjeonglee@ucla.edu (Y. Lee).

https://doi.org/10.1016/j.specom.2021.05.001
Received 7 July 2020; Received in revised form 13 March 2021; Accepted 3 May 2021
Available online 8 May 2021

speaking rate and importantly found that the magnitude of that variability differs from speaker to speaker, as was also found in Perkell et al. (2008). Recognition of such within-speaker variability has led to the view that a speaker's representation of a unit is a distribution over the quantitative values that goal parameters can take on. While early "window" models of such representations (Keating 1990, 1996; Guenther 1994, 1995; Byrd 1996; Saltzman and Byrd 2000) entertained the possibility of essentially rectangular distributions, i.e., windows as acceptable *ranges* of parameter values, other work has hypothesized Gaussian-type distributions (e.g., Roon and Gafos 2016; Villacort, Perkell and Guenther, 2007) or arbitrary distributions built up from category exemplars (e.g., Johnson 1997; Pierrehumbert 2001). One possible type of evidence that speakers' representations have such Gaussian character is that speakers' real-time behavior is sensitive to the prototypicality of the token that they are producing, i.e., how close it is to the center of the distribution. For example, Niziolek Nagarajan and Houde (2013) report that tokens of a vowel that are produced relatively far from a speaker's mean target early in the vowel's production tend to be corrected towards the mean later in the vowel's production (though this could also be compatible with rectangular windows, if the peripheral tokens fall completely outside that window). The variability in phonological representations also affords the speaker some flexibility in how the same unit is produced in different contexts. Harper (2020) has shown that the within-speaker variability observed within a phonetic context is also predictive of how much a speaker's production of that unit shifts across different contexts. In a shadowing study, Lewandowski and Nygaard (2018) revealed a potential contribution of shadowers' articulatory flexibility measured from vowel dispersion to the extent of individual vocal alignment behavior.

A second key component underlying the capacity for real-time accommodation lies in the *directness of sensory-motor correspondence* in the human orofacial system; sensory activity resulting from orofacial movements (self-generated or generated by another) potentially and transparently engages the corresponding motor activity. This transparent correspondence has been argued to result from a "common currency" between listeners and producers (Goldstein and Fowler 2003) that simultaneously represents both sensory and motor information in terms of the properties of biologically significant actions in the world. In speech (which engages the orofacial system), the correspondence is a key component in satisfying what is known as the "parity requirement" that transmitted and received messages be the same (e.g., Liberman and Whalen 2000) or sufficiently equivalent (Goldstein and Fowler 2003), that a listener successfully both recognize and identify language forms (Fowler and Magnuson 2012). One source of evidence for this correspondence is the ability of infants to imitate facial gestures. Meltzoff and Moore (1977, 1997) discovered that newborns (the youngest 42 min old) can imitate the facial gestures produced by a caregiver. The baby cannot see its own face nor "feel" the face of the caregiver, and therefore some robust (and presumably innate) sensorimotor correspondence is needed to account for this, whether by means of "common currency" or some other mechanism.

A third key cognitive component leading to accommodation behavior is the socially induced *pressure for an individual to act similarly* to others (Giles, Coupland and Coupland, 1991; cf. Pickering and Garrod 2004). It has been argued (Goldstein and Fowler 2003) that this systematic tendency for individuals to attune their behavior to one another can, when supported by a transparent sensory-motor correspondence, play a key role in the formation and stabilization of phonological categories that are shared among members of a speech community. Illustrations of how this might be so have been offered via simple simulation models showing the emergence of categories along some continuous articulatory dimension(s) in a system of computational agents (e.g., Goldstein 2003; Goldstein et al., 2008; cf. Oudeyer 2006, for a similar approach). In Goldstein's work, at the beginning of the simulation two agents produce random values of a continuum (uniform probability distribution). On each further production trial, both agents emit a value at random from their probability distribution. If the agents' two values match within some noise threshold (determined in part by the nature of the degree of uncertainty of the sensory-motor correspondence), then both agents increase slightly the probability of producing the step values they just produced. This is repeated over many trials, and the probability distributions continually evolve. Ultimately, both agents develop one or more narrow Gaussian-like distributions centered on one of the continuum steps, with the step location values of the two agents matching each other within the noise threshold. In phonological terms, this provides for the creation of the values that could be produced by structured phonological categories. These simulations are a conceptual model of how categories along a continuum could emerge without beginning with any structure at all; they emerge from sensory-motor correspondence combined with the social intention to behave similarly to other individuals we interact with.

2. Attunement model

2.1. Modeling foundations

It is possible to use the same type of computational model to simulate how online accommodation can emerge from the flexible adaptability hypothesized to exist due to the three underlying cognitive capacities of variability, direct correspondence, and pressure to act similarly. Such a simulation approach adopts two simple computational foundations—a random choice of values from each agent's continuum on every trial and an increasing probability of a value that is produced on a trial in which the two agents match. This type of model makes several predictions about observable features of accommodation, and these can be tested not only with the results of performing the simulations but also with real behavioral accommodation data.

In such a simulation model of two interacting interlocutors (agents), instead of beginning with each agent randomly choosing values along a continuum as in the simulations described above, both 'conversing' agents can begin with Gaussian probability distributions centered on different steps of a continuum that represents some control parameter of the speakers' performance before they interact with another. The centers (μ) of the two agents' distributions along the continuum can be manipulated across different simulations, as can the standard deviation (σ) of the each of two agents' starting distributions. In the course of a simulation, convergence of the agents' distributions is expected under a subset of initial conditions.

The structure of this very simple model leads to three predictions about observed convergence. (1) Convergence will occur on a simulation only if the μ and σ values are such that there is a non-zero probability that a randomly produced match will occur. If the μ values are far apart and the σ values are small, no matches may ever occur by chance, and therefore no convergence is predicted to occur. (2) The more flexible agent—the agent with a higher value of σ—will exhibit a greater shift of mode in converging. This is because the more flexible agent is more likely than the other agent to produce values far from their μ value that will result in matches, and these will eventually push their μ value a greater distance. (3) That said, both agents may show some reduction in variability of the values they produce after convergence as compared to before, because the way matching is rewarded will tend to increase the maximum probability associated with mode values, thereby reducing σ. Nevertheless, the change is expected to be greater for the more flexible agent, as their distribution comes to resemble the distribution of the less flexible agent.

It is unclear if convergence and divergence involve separate or (inter) dependent processes/mechanisms. Thus, our conceptual model is constrained to address predicting accommodation behaviors specifically for interactions exhibiting convergence, not the cases of divergence or cases of no observable accommodation.

2.2. Simulation method

The simple attunement simulation described above can be specified with these further details in our implementation. At the beginning of each simulation, two agents, A_1 and A_2, are assigned normal probability density distributions associated with selecting (producing) each of the steps of a continuum representing some articulatory or acoustic parameter of speech. Here 70 steps (−35 to 34) are used in the simulations. Each agent's distribution is defined by choosing a distribution mean value (μ_1 and μ_2) and a standard deviation (σ_1 and σ_2). Fig. 1a shows an example of initial distributions with $\mu_1 = -3$, $\sigma_1 = 2$, and $\mu_2 = 3$, $\sigma_2 = 4$. On each iteration of the simulation, each agent selects a value from her particular distribution. If the two values match within a noise criterion that represents uncertainty in the sensory-motor correspondence, ($|x_1 - x_2| < noise$), each agent increases the probability associated with the value that she just produced. The noise criterion used is 3 continuum steps, and rule updating the probability is: $P_{n+1} = P_n + 0.001 (1 - P_n)$. Iterations are divided into epochs of 250 iterations, and epochs continue until *both* of the agents stabilize their productions, where stabilization is a change in the mean production from one iteration to the next that is below a *criterion: mean* ($|\bar{x}_{i+1} - \bar{x}_i| < criterion$, where *criterion* is 2 continuum steps. If the stabilization criterion is not achieved by iteration 50,000, the simulation ends. The mean values selected by the two agents at the final epoch are compared, and the simulation is considered to show convergence if $|\bar{x}_1 - \bar{x}_2| < noise$. Fig. 1b shows the probability function at the final iteration of the converged simulation for the initial condition in Fig. 1a. Fig. 1c tracks the mean value ($\bar{x}$) of each agent across epochs, and Fig. 1d does the same for the variance (*Var(x)*) of each agent.

Initial conditions for the simulations compared three values of the distance between μ_1 and μ_2: $\mu_2 - \mu_1 = 6$, $\mu_2 - \mu_1 = 12$, $\mu_2 - \mu_1 = 18$, symmetrically placed around 0 (e.g., $\mu_1 = -3$, $\mu_2 = 3$). The standard deviations of the agents (σ_1 and σ_2) used all combinations of three values: 2, 4, and 8. Combining the σ combinations with all μ values gives $3 \times 3 \times 3 = 27$ distinct simulation conditions. Each condition was simulated 25 times, yielding a total of 675 simulations.

2.3. Simulation results

Of the 675 simulations, 34 simulations (5%) failed to converge, with the majority (32/34) failing to reach the stabilization criterion. Consistent with prediction (1), the distance between the μ values (agent means) was maximal ($|\mu_2 - \mu_1| = 18$) on 26 of these convergence failures and was large for the other six ($|\mu_2 - \mu_1| = 12$), while the σ (standard deviation) values were minimal: in 31 of these convergence failures $\sigma_1 = 2$ and $\sigma_2 = 2$ and in the other case $\sigma_1 = 4$ and $\sigma_2 = 2$. In six of the 30 non-converged simulations, both agents shifted their mean values towards each other by the final (50 000th) iteration in the direction of convergence—i.e., a reduced distance between agent means. Five of these approximation cases were where $|\mu_2 - \mu_1| = 12$ and $\sigma_1 = 2$ and $\sigma_2 = 2$, and one case was where $|\mu_2 - \mu_1| = 18$ and $\sigma_1 = 4$ and $\sigma_2 = 2$. In the other 24 cases, in which $|\mu_2 - \mu_1| = 18$ and $\sigma_1 = 2$ and $\sigma_2 = 2$, the mean values of the two agents remained unshifted and far from one another through the iterations. An additional four simulations that converged failed to reach the stabilization criterion. Further analyses are based on the 639 simulations that both reached criterion and converged.

For the 95% of simulations that stabilized and converged, for each agent the shift in mean value from the initial epoch to the final was calculated: $Shift_i = |\bar{x}_{final(i)} - \bar{x}_{initial(i)}|$. In 446 converged simulations, one agent had a higher initial σ than the other. *In every instance, the more variable agent shows a greater shift.* In other words, consistent with prediction (2), the agent with a higher baseline variability is the 'converger.'

In the remaining 193 converged simulations in which $\sigma_1 = \sigma_2$, there is no preference for which agent shows the greater shift—95 times it was A_1 and 98 times it was A_2. In addition, the relative contribution of each agent to convergence is quite variable across individual simulations, as can be seen in the scatterplot in Fig. 2, which plots the *Shift* of A_1 (shift in mean value) against the *Shift* of A_2 for all the simulations where $\sigma_1 = \sigma_2$. The three clusters of points correspond to the $|\mu_2 - \mu_1|$ conditions: 6, 12 and 18 units. Each cluster exhibits a range of relative contribution of the agents' shifts to the convergence. In some simulations, only one agent shifts a substantial amount (peripheral points in the cluster), while in other simulations, the two agents shift nearly equally (the central points in a cluster), and all intermediate possibilities are attested. Since in all cases the final values of μ_2 and μ_1 are converged (close to equal), the change of the two agents must sum to the initial $|\mu_2 - \mu_1|$ value in each condition ± the noise parameter. Thus, there is a negative correlation between the change in the two agents in each condition ($r = 0.23$, $p < 0.005$). This is quite different from the simulations in which initial σ differs between the two agents. In these 446 converged simulations, shift of the flexible agent is almost entirely responsible for the convergence: *mean(Shift)* ± standard error for the flexible agent (higher initial σ) is 10.29 ± 0.75 (A_1, when $\sigma_1 > \sigma_2$) and 10.49 ± 0.77 (A_2, when $\sigma_1 < \sigma_2$) while *mean(Shift)* for the less flexible one is only 1.0 ± 0.08 (A_1, when $\sigma_1 < \sigma_2$) and 1.2 ± 0.1 (A_2, when $\sigma_1 > \sigma_2$).

Finally, the difference between the variance for each agent during the final epoch *Var(x)* and the agent's baseline variance (σ^2 of the initial probability distribution) was calculated (*Var(x) - baseline*). Fig. 3 presents violin plots (with overlaid box plots) of *Var(x) - baseline* for A_1 and A_2 separately for three different simulation conditions: $\sigma_1 > \sigma_2$, $\sigma_1 = \sigma_2$, $\sigma_1 < \sigma_2$. When $\sigma_1 > \sigma_2$ there is a substantial reduction in *Var(x)* from its baseline value (*mean* (*Var(x) - baseline*) = −43 ± 3.3). The reduction in *Var(x)* is smaller, but still present, in the other conditions (mean reduction is −28 ± 2.8 when $\sigma_1 = \sigma_2$ and is −5.9 ± 0.53 when $\sigma_1 < \sigma_2$). Thus, consistent with prediction (3), when A_1 begins with higher variability than A_2, A_1 shows a substantial decrease in variability during the convergence process. In other conditions, there was also some decrease in variability, but much smaller, again as predicted. Comparable patterns are observed with A_2; again, when A_2 is more variable than A_1, A_2 shows the largest decrease in variability during the convergence process.

3. Experimental assessment

We turn next to an experimental dataset that serves to assess the hypothesis that structured variation can serve as an index of individual speaker adaptability underlying convergence behavior in speech accommodation, which we have illustrated in an elementary way in the modeling above. In the present experimental study, we examine 1) how the variability structure in acoustic and kinematic properties of individuals' speech may differ when they are engaged in a cooperative dyadic speech activity as compared to a similar solo speech activity, and 2) how variability differences between partners in a dyadic interaction relate to the speech convergence phenomena that they exhibit. We specifically examine whether and how the individual speaker variability may serve as an indicator of which speaker of a dyad converges to their partner.

To address these questions, we explore a dataset reported in a previous accommodation study (Lee et al., 2018) that collected acoustic and articulatory kinematic data from pairs of facing speakers jointly participating in a maze navigation task as well as from each speaker completing the same task individually prior to the joint task. The data from three pairs of speakers who demonstrated convergence in Lee et al. (2018), out of the original four dyads, are investigated in the present study.[1] Fig. 4 shows the results of significant convergence patterns observed in Lee et al. (2018). Speakers' production behaviors are examined in solo speech (SOLO) before the conversational interaction and then during that interaction. The 'converger' was defined as the member

[1] The fourth dyad showed instances of 'divergence' behavior, which is not further discussed in the present follow-up study.

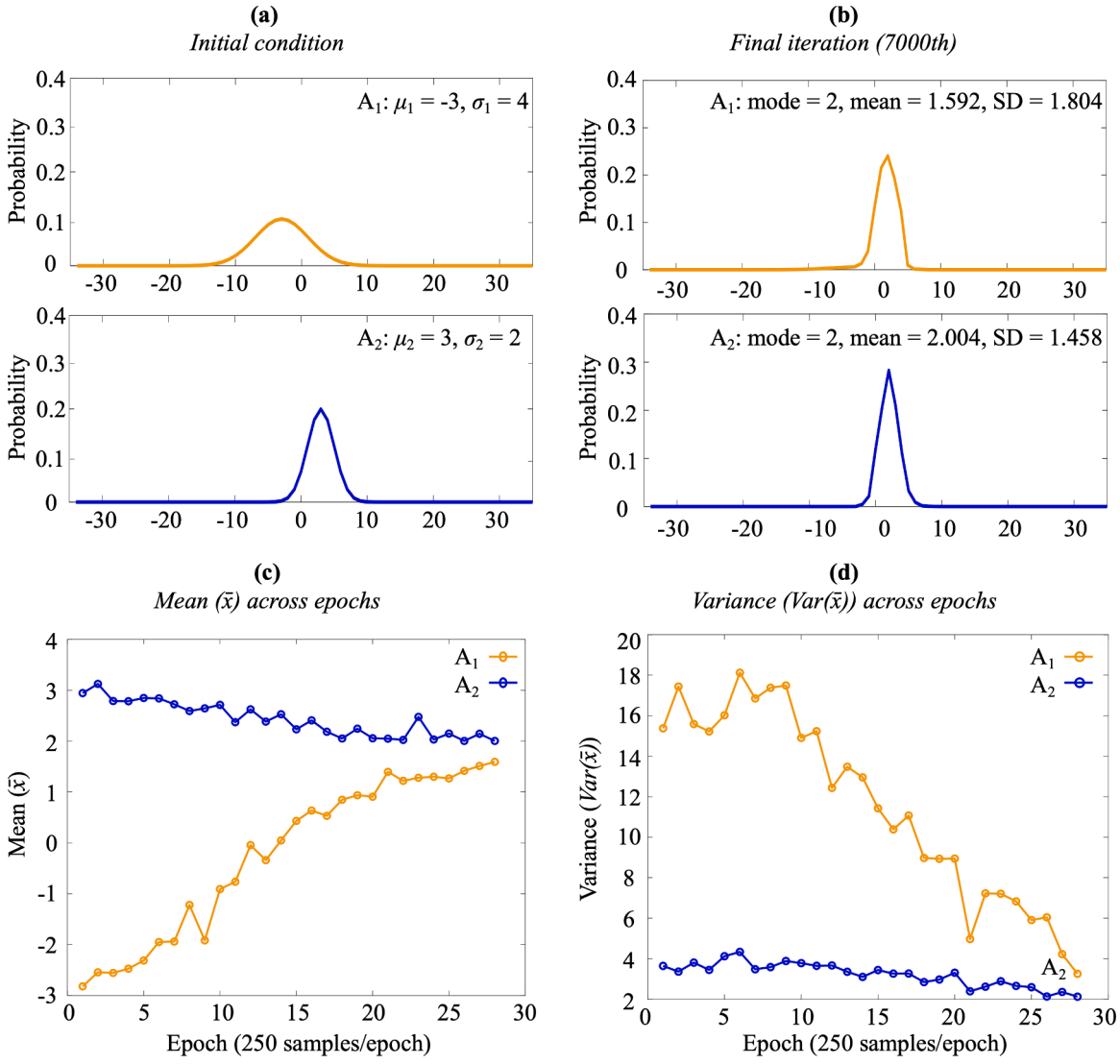

Fig. 1. Simulation method, illustrative example. (a) probability function of initial distributions for: $\mu_1 = -3$, $\sigma_1 = 4$, and $\mu_2 = 3$, $\sigma_2 = 2$. x axis represents the continuum steps; (b) probability function at the final iteration of the converged simulation for the initial condition in (a); (c) tracking of the mean value ($\bar{x}$) of each agent across epochs; (d) tracking of the variance (Var($\bar{x}$)) of each agent across epochs.

of the dyad that shows a larger shift in mean values between the SOLO condition and the INTERACTION condition (compare the Δmedian of mean values of the two dyad members in Fig. 4). Speakers converge during the interactive task (INTERACTION) in various measures: sentence duration, the stiffness control parameter underlying the tongue movement (indexed by time-to-peak-velocity), and intonational measures. While in some cases both speakers of the dyad become more similar to one another, in all cases *one* speaker of the dyad is particularly malleable, producing the large preponderance of the convergence or accommodation toward their partner relative to their own prior solo speech; these convergers are denoted using orange in the figures.

3.1. Methods

3.1.1. Samples and stimuli

This study used simultaneously collected audio and articulatory data reported in Lee et al. (2018). The full data set and a guideline to its organization are freely available via a public repository at https://doi.org/10.5281/zenodo.1119284.

Participants were adult native speakers of American English with typical speech and hearing (mean age = 25). They were paired into dyads of the same sex: one male dyad denoted as Dyad S1-S2 and three female dyads denoted as Dyads S3-S4, S5-S6 and S7-S8—making a total of four dyads. Dyad members were not previously familiar with one another. Note that this study examines only the convergence patterns (Fig. 4), so the diverging dyad, Dyad S5-S6, was excluded from the analysis.

Details of the experiment setup, data acquisition, and post-processing are fully described in Lee et al. (2018). Briefly, the synchronized audio and articulatory data of participants were recorded in a sound-insulated room. Each speaker was seated with a table-top microphone in front of them and beside an electromagnetic articulography (EMA) system, facing their dyad partner at a distance of about three meters. The kinematic data analyzed in the present study was drawn from the movement tracking of an EMA sensor coil placed on the tongue tip.

Speakers participated in speech tasks presented on their computer monitors (see Table 1): a sentence reading task (solo pretest, not analyzed), a maze navigation task independently completed (solo speech #1), a maze task cooperatively completed with one another (interactive speech), followed by another independently completed maze task (solo speech #2, previously analyzed but not included in this

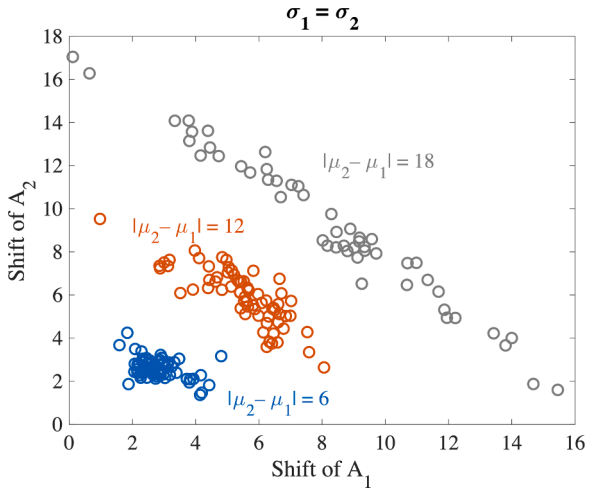

Fig. 2. Simulation results. Scatterplot of Shift values of A_1 against Shift values of A_2 for the 193 converged simulations in the condition $\sigma_1 = \sigma_2$, where $Shift_i = |\bar{x}_{final(i)} - \bar{x}_{initial(i)}|$. Each point represents the results from one of the simulations.

new analysis) and finally a reading task (solo posttest, not analyzed). During an 'individual' task, an opaque screen was placed to block the line of sight between the dyad members, and headphones with music playing were given to a dyad member who was not performing the task to prevent them from hearing the other dyad member's speech.

The carrier sentence used in the tasks had phrase-medial ("beside" or "between") and phrase-final ("signs" or "lights") target words, resulting in four combinations:

"And then you go ____ [beside/between] the next two ____ [signs/lights]."

For the maze navigation tasks, the mazes were designed to have a balanced occurrence of the target word pairs ("between/beside" and "lights/signs"), and each maze landmark icon was either two road signs or two traffic lights. At each landmark, the speakers were asked to use the frame sentence and appropriate combinations of target words to describe that landmark.

Fig. 5 is an example maze image presented to a participant in the interactive task. For each unique maze, the two dyad members saw different views of the same maze, differing only in the locations of solid and dotted blue lines. The dyad members were asked to navigate only the sections indicated by solid lines and to turn over the floor to their partner to navigate the other sections with dotted lines. (The mazes in solo speech had only solid lines.) At each turn, each dyad member navigated either one or two landmarks in a row.

For the solo conditions, participants were presented with eight unique individual mazes. For the interactive condition, in which both dyad members participate, eighteen different versions of cooperative mazes were repeated twice in random order, yielding a total of 36 mazes.

The target words analyzed here are the phrase-medial "beside" and phrase-final "signs," which share comparable articulatory trajectories: the diphthong [ɑɪ] in each [bəsɑɪd] and [sɑɪnz] is followed by tongue tip constriction and release.

3.1.2. Data analysis

Among the acoustic and articulatory variables quantitatively analyzed in Lee et al. (2018), the present study considers specifically only those variables that showed significant between-speaker convergence effects. In most cases, the convergence was accomplished by just one of the two speakers in a dyad, with that speaker becoming more like their partner (i.e., a reduced phonetic distance between the two dyad members). In Lee et al. (2018), significant convergence was determined to have occurred when a measure became more alike in the interactive condition, compared to its behavior in the solo condition before interaction (solo speech #1). As shown in Fig. 4, the measures in Lee et al. (2018) that were seen to have converged included acoustic sentence duration (Dyads S1-S2, S3-S4, and S7-S8), utterance-final f0 maximum for the H% boundary tones (Dyad S3-S4), and time-to-peak-velocity (TPV) for the tongue tip closure (Dyad S7-S8) or release (Dyad S1-S2). Note that TPV is an index of gestural stiffness control parameter. The reader is referred to Lee et al. (2018) for further details on measurements and the breadth of findings in the earlier study.

The present study turns its focus to the structured variability observed in this dataset for the maze condition (SOLO vs. INTERACTION). For a given measure of variability, there were 8 mazes for the SOLO maze condition before interaction and 34–41 mazes for the INTERACTION condition, each of which with ~4 observations. To quantify the amount of variability in a speaker's speech over the course of experimental trials, we use a moving coefficient of variation (moving CoV = moving σ / moving μ * 100). For each parameter from each speaker in each maze condition, a rolling calculation window of eight observations was set,

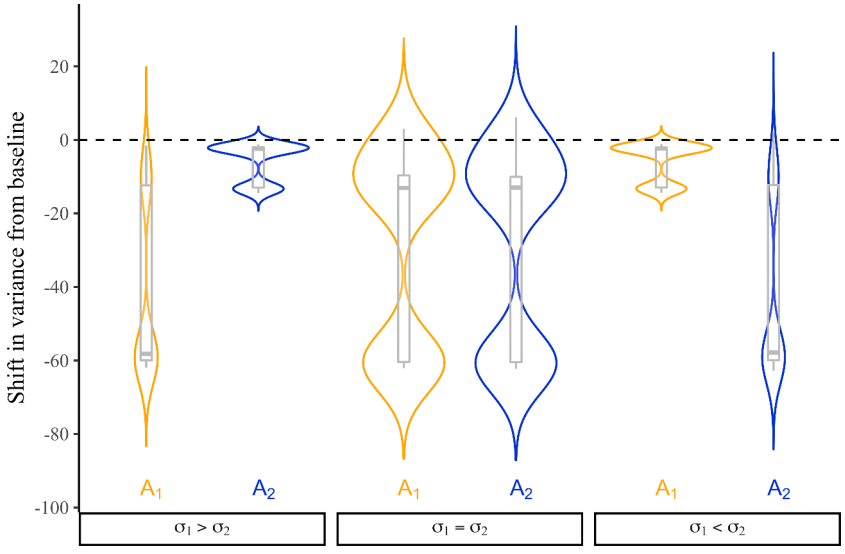

Fig. 3. Simulation results. Violin plots with overlaid box plots of Var(x) - baseline for A_1 and A_2 for three simulation conditions: $\sigma_1 > \sigma_2$, $\sigma_1 = \sigma_2$, $\sigma_1 < \sigma_2$. Black dashed horizontal line at 0 = agent's baseline variance.

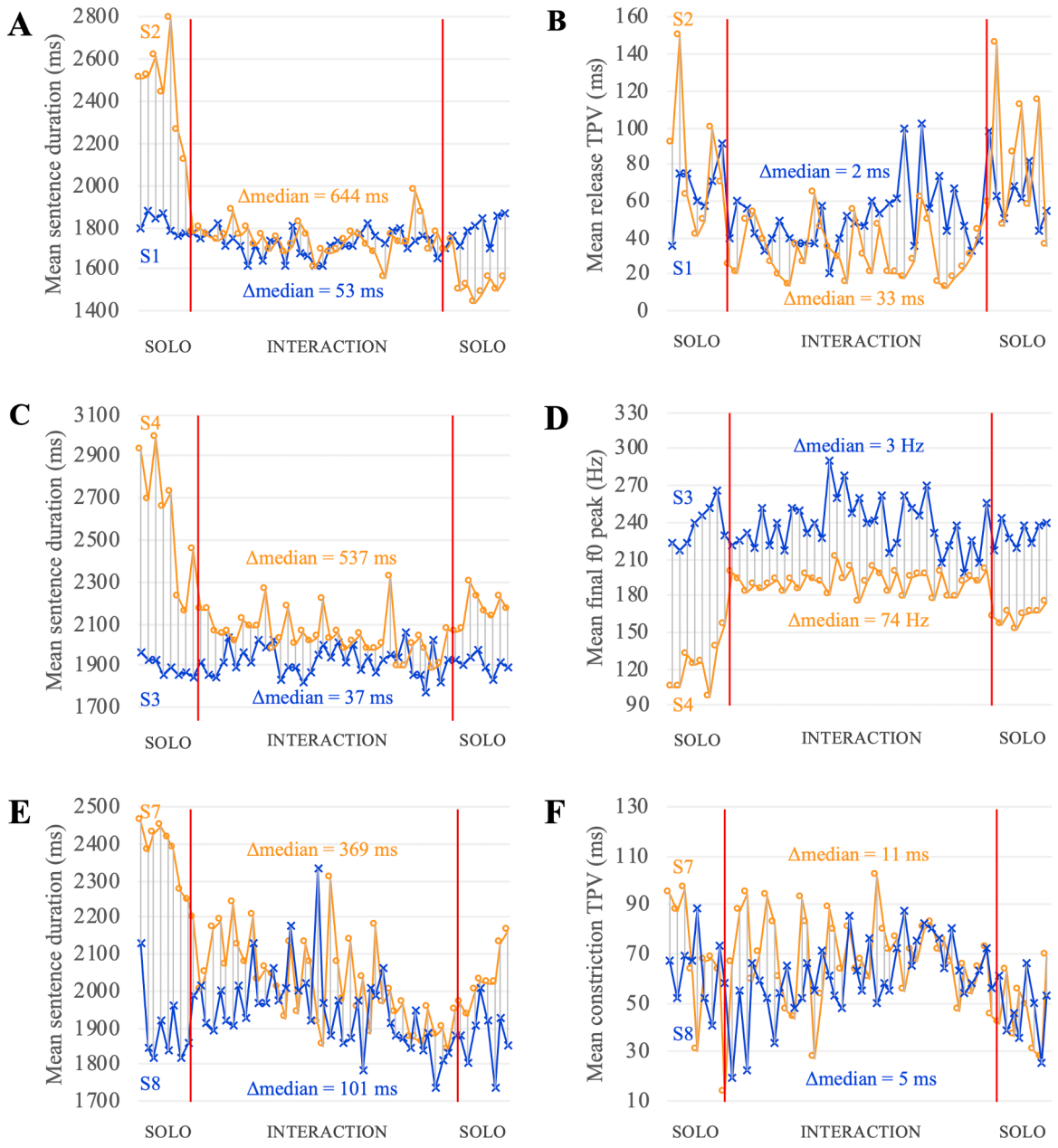

Fig. 4. Measures with significant convergence over the course of maze trials. TPV = time-to-peak-velocity. Each data point along the x axis represents a mean value of parameter for each maze trial. Δmedian = absolute difference in median of the mean values between the INTERACTION condition and SOLO condition prior to interaction. Adapted from Lee et al. (2018).

substituting two new values for the two oldest values with each advance of the window.

This study assesses the changes in variability from when speakers were speaking solo to when they were interacting with one another for each maze. The following measures are calculated:

- Baseline CoV: the individual variability inherent to an individual speaker's solo speech.
  - Baseline CoV was assessed by comparing moving CoV values of the two dyad members in their solo maze conditions (before the cooperative maze task).
- Absolute difference in CoV: |ΔCoV|, where ΔCoV = CoV during interaction - mean baseline CoV.
  - For quantifying changes in variability structure, for each speaker, difference values (Δs) were calculated by subtracting the mean value of baseline CoVs (a constant value) from each of the CoV values during interaction, as calculated by the moving window described above.
  - A ΔCoV value indicates either a decrease (-Δ) or increase (+Δ) in individual variation during their interactive speech as compared to their solo speech. To assess the magnitude of shifts in CoV associated with the interactive maze condition, absolute values of ΔCoV (i.e., |ΔCoV|) for a given measure are compared.

The differences in the paired values of baseline CoV and |ΔCoV| in the acoustic and kinematic measures of the two members for a given dyad are statistically evaluated using the non-parametric sign test, with p-values ≤ 0.05 considered significant.

Table 1
Task presentation order for Dyad A-B. Conditions in gray cells were not analyzed in this study.

Condition	Participant	Task
pretest	Speaker A	individual sentence reading task
pretest	Speaker B	individual sentence reading task
solo speech #1	Speaker A	individual maze task
solo speech #1	Speaker B	individual maze task
Interaction	Speaker A & B	cooperative maze task
solo speech #2	Speaker A	individual maze task
solo speech #2	Speaker B	individual maze task
posttest	Speaker A	individual sentence reading task
posttest	Speaker B	individual sentence reading task

3.2. Experiment results

3.2.1. Distributions in solo versus interactive speech

Fig. 6 shows full distribution patterns of each significant accommodation variable from each speaker of the three converging dyads during solo speech and interactive speech. The reader is also referred to the model simulation results in Fig. 1a (initial distributions) and 1b (the converged simulation). For all three dyads, during their solo speech the malleable member of a dyad (orange) always shows a wider density curve (i.e., a higher value of σ) compared to the other member (blue). During the interactive trials, the distribution belonging to the originally more variable member resembles that of their dyad partner. As shown in shifts in median values (vertical dashed lines) and changes in widths of the curve from solo to interactive trials, the member with higher baseline variability is always the one who adapts more to their partner and exhibits a significantly reduced value of σ. We additionally note that the baseline variability between the members of Dyad S7-S8 seems comparable to another for the tongue tip closure time-to-peak-velocity measure (Fig. 6F). These patterns are statistically confirmed in the detailed analysis of time-varying changes in variability presented in the following subsections.

In solo speech, five of six cases show that at least one production value of each dyad member falls within the range of the values of the other member, exhibiting overlapping distributions between paired speakers. In these cases, speakers converge to nearly identical production values, with similar means and largely overlapping distributions. The one exception is the phrase-final f0 peak measure from Dyad S3-S4 (Fig. 6D), in which, unlike the others, the baseline distributions of f0 values from Speaker S3 and Speaker S4 do not overlap. Despite the lack of a precise match in median values during the interactive trials, the 'converger' (Speaker S4) still shows a significant shift in their production values towards their partner's f0 values such that the two speakers do exhibit overlapping distributions of f0.

3.2.2. Baseline coefficient of variation (CoV) in solo speech

For all three converging dyads, the baseline CoV values in solo speech are always greater for the dyad member who then went on to converge in the interactive measure(s) as compared to the baseline CoV values for the other dyad member who was found to be less malleable during the interactive speech. In all cases below, we report means ± standard deviations of the CoV measure.

For Dyad S1-S2, the baseline CoV values for sentence duration are consistently greater for Speaker S2, the more malleable member of the dyad, than for Speaker S1 (S1: 4.2 ± 1.5, S2: 15.1 ± 7.8, $z = 3.32$, $p < 0.001$). The phrase-final tongue tip release time-to-peak-velocity (TPV) has a similar pattern with Speaker S2 again exhibiting more variability (67.0 ± 16.9) than Speaker S1 (48.4 ± 19.7), though this difference is not statistically significant ($z = 1.51$, $p = 0.13$) due to the three instances (out of 11) that show the opposite pattern (S2 < S1).

During interaction, Dyad S3-S4 converged in sentence duration and utterance-final f0 peak measures. For this dyad, the baseline CoV is significantly greater for Speaker S4, and as predicted it was S4 who drove convergence in both measures, as compared to their less malleable partner, Speaker S3 (sentence duration, S3: 6.2 ± 2.6, S4: 12.3 ± 5.5; f0, S3: 7.2 ± 2.6, S4: 21.2 ± 1.8; all $p < 0.005$).

For the final converging dyad, Dyad S7-S8, in which Speaker S7 is the malleable member, Speaker S7 is slightly more variable than Speaker S8 in sentence duration (S8: 6.3 ± 2.6, S7: 7.2 ± 3.4) and tongue tip constriction TPV (S8: 41.0 ± 8.6, S7: 48.1 ± 17.7), but neither difference is significant (sentence duration: $z = 0.19$, $p = 0.85$; TPV: $z = 0.58$, $p = 0.56$).

3.2.3. Difference CoV values: |ΔCoV|

Across dyads, a greater change in the CoV values from solo speech to interactive speech is indicated by higher |ΔCoV| values. As predicted, the higher difference value is always associated with the more malleable dyad member. That being said, we observe mixed results with respect to the direction of the condition-dependent shift in CoV (+/-ΔCoV). These results can be visually confirmed in Fig. 7, in which the less malleable speaker within a dyad is marked with blue x symbols, and the dyad member driving convergence during interaction is indicated with orange circles. The black bold lines placed horizontally at 0 refer to

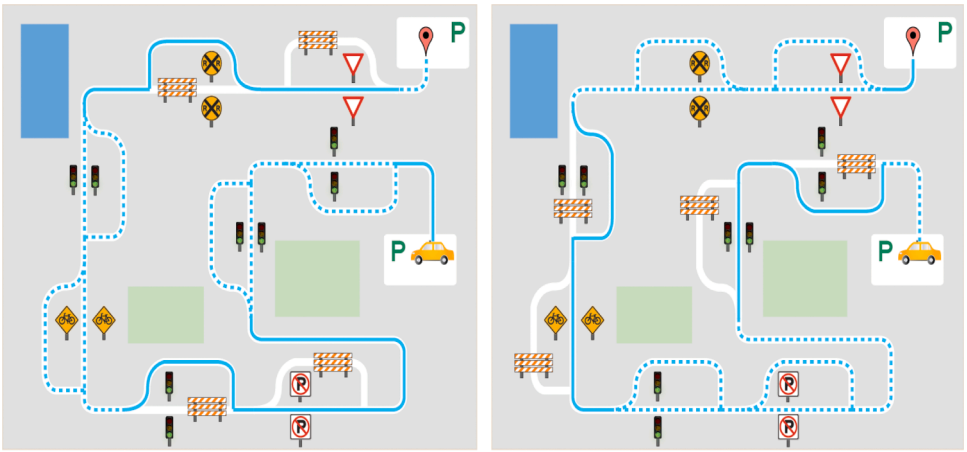

Fig. 5. Example cooperative maze trial. Adapted from Lee et al. (2018).

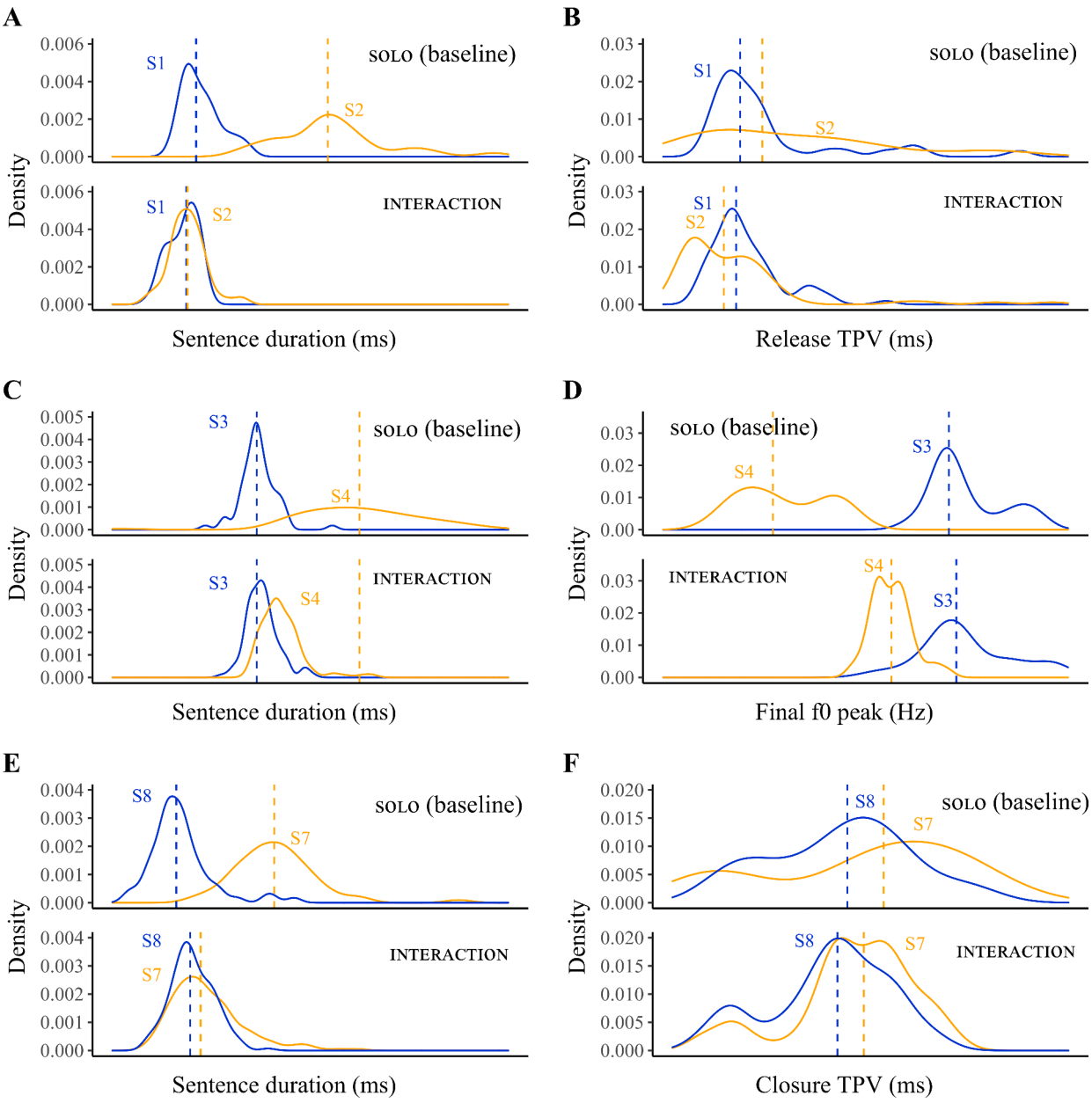

Fig. 6. Density plots of accommodation variables in converging dyads in solo (top plots) versus interaction (bottom plots) tasks: Dyad S1-S2 (A,B), Dyad S3-S4 (C,D), & Dyad S7-S8 (E,F). A vertical dashed line indicates the median value of a distribution. Orange indicates the 'converger' within the dyad.

speakers' baseline CoVs before interaction. Thus, regardless of its direction, the distance from 0 to each data point indicates the magnitude of shifts in CoV associated with the interactive condition (= |ΔCoV|).

For Dyad S1-S2's sentence duration (Fig. 7A) and final tongue tip release TPV (Fig. 7B) measures, the magnitude of shift in CoV from solo speech to interactive speech is greater for Speaker S2 (the 'converger') than for Speaker S1 (mean sentence duration |ΔCoV|, S1: 0.65 ± 0.48, S2: 10.44 ± 1.19, $z = 5.39$, $p < 0.001$; mean release TPV, S1: 13.46 ± 6.77, S2: 24.7 ± 18.69, $z = 2.47$, $p < 0.05$). As shown in the top figure panels, ΔCoV values for the malleable speaker (S2) are farther away from zero, the baseline CoV. In contrast, ΔCoV values for the less malleable partner (S1) do not deviate much from their baseline. In addition, Speaker S2 consistently shows negative ΔCoV values in sentence duration, while exhibiting largely fluctuating ΔCoV values in the release TPV measure.

For Dyad S3-S4, a robust between-speaker difference in |ΔCoV| is observed for both sentence duration (Fig. 7C, S3: 1.82 ± 1.09, S4: 6.3 ± 2.94, $z = 6.06$, $p < 0.05$) and f0 (Fig. 7D, S3: 3.92 ± 2.42, S4: 14.92 ± 1.82, $z = 5.66$, $p < 0.001$). Again, the 'converger' (S4) shows a greater distance between CoV during interaction and their solo speech baseline CoV than does their less malleable partner (S3), in both temporal and intonational variables. Across measures, Speaker S4 shows negative ΔCoV values (excepting the two positive values near zero in sentence duration), whereas Speaker S3 shows ΔCoV values fluctuating around the baseline.

Lastly, Dyad S7-S8 patterns similarly to the other converging dyads, again, showing that the 'converger' of the dyad (S7) exhibits larger fluctuations in ΔCoV. However, the difference between speakers is not significant for either sentence duration (Fig. 7E, S7: 2.46 ± 2.14, S8: 1.69 ± 1.08, $z = 0.38$, $p = 0.71$) or TPV (Fig. 7F, S7: 18.55 ± 10.92, S8: 12.25 ± 8.57, $z = 1$, $p = 0.31$). For sentence duration (Fig. 7E), the ΔCoV values for the malleable speaker (S7) fluctuate in both directions from the baseline, whereas the values for the less malleable speaker (S8) mostly occupy the region below zero during interaction. For the

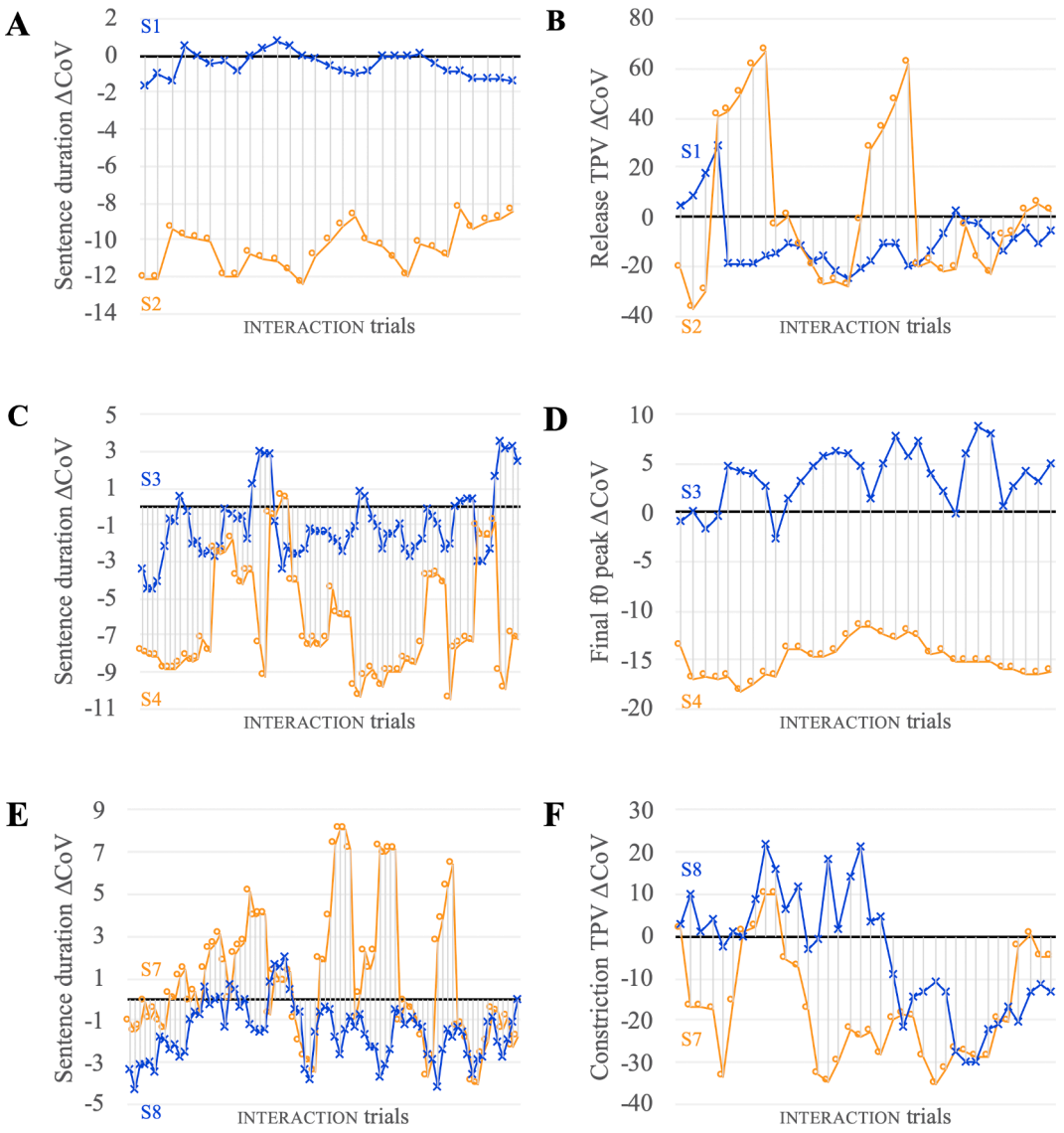

Fig. 7. ΔCoV of accommodation variables over the INTERACTION task trials (x axis) in converging dyads (i.e., the interval between the red vertical lines in Fig. 4): Dyad S1-S2 (A,B), Dyad S3-S4 (C,D), & Dyad S7-S8 (E,F). ΔCoV = change in coefficient of variation from solo speech to interactive speech. Black bold horizontal line at 0 = speaker's baseline CoV measured in solo speech. Orange indicates 'converger' within the dyad.

constriction TPV measure (Fig. 7F), both speakers show various ΔCoV values throughout the interaction trials, with the overall distance from the baseline slightly larger for the malleable speaker than for their partner.

Table 2
A summary table.

	Significant measures of convergence	baseline CoV	|ΔCoV|
Dyad S1-S2	Sentence duration	S1 < S2	S1 < S2
	Phrase-final tongue tip release TPV	S1 < S2	S1 < S2
Dyad S3-S4	Sentence duration	S3 < S4	S3 < S4
	Phrase-final f0 peak	S3 < S4	S3 < S4
Dyad S7-S8	Sentence duration	S8 < S7	S8 < S7
	Phrase-final tongue tip constriction TPV	S8 < S7	S8 < S7

An underlined subject (S#) indicates the speaker driving convergence within a dyad; gray cells indicate $p > 0.05$. [CoV = coefficient of variation; TPV = time-to-peak-velocity].

3.2.4. *Summary of experimental assessment*

Table 2 synthesizes these results. In the table, underlining indicates the malleable dyad member who converged during interaction in the indicated measure: Speakers S2, S4 and S7.

In sum, the variability inherent to an individual speaker's solo speech—assessed by baseline CoV—is consistently greater for the malleable dyad member than for the less malleable dyad member (significantly so in three of six cases). This direction of difference is never the reverse.

Changes in variability structure, indexed by the absolute difference CoV values from individuals' solo speech to interactive speech (|ΔCoV|), are consistently greater for the 'convergers' than for their dyad partners (significantly so in four of six cases). The direction of difference is never the reverse.

Additionally, we note that when any baseline production values from the two dyad members happen to fall within each other's range (a non-zero probability), the convergence pattern during interaction shows near-perfect matches between the adapted values of the dyad members. In one case in which baseline distributions of the dyad members do not overlap (phrase-final f0 peak of Dyad S3-S4), shifts in f0 values do

reduce the inter-partner phonetic distance, but an imperfect match between the interspeaker distributions is observed—the members' median values remain far from one another.

Finally, as shown in Fig. 7, the more flexible speakers (S2, S4 & S7) generally (in four of six cases) show reduced (negative) CoV values in their production after convergence has occurred, as compared to speaking solo. In the two exceptions to this pattern (S2's TPV [upper right] and S7's sentence duration [bottom left]), the malleable speakers fluctuate relative to their baselines. In no case is there an increase in the variability of these speakers.

4. Discussion of structured variability as an index of individual adaptability

Conversing speakers often attune their speech behavior to one another. Much evidence for speech accommodation exists, yet we know relatively little about the cognitive capacities that may lead to online adaptations in speakers' production. This study expounds how underlying components of the capacity for accommodation may allow two speakers engaged in a dynamic spoken interaction to stabilize their speech productions. We present a simple attunement model that serves as a conceptual basis for constructing predictions about accommodation behaviors and test the model's predictions by examining a dataset of pairs of conversing speakers exhibiting convergence behaviors.

Both model simulation and experimental results support our overarching hypothesis that structured variation may reveal individual speaker adaptability that underlies convergence behavior in speech accommodation. Real-time accommodation is observed in our simple model of two computational agents (emulating two conversing interlocutors) that instantiates the hypothesized key cognitive components underlying convergence. These components are (i) individual agents' adaptability that springs from their natural variability in producing a phonetic unit, (ii) transparent sensory-motor correspondence, and (iii) 'social' pressure to behave similarly.[2] In the case of two agents who produce their representative performance values not too far from one another along a (phonetic) continuum, matching of their mean values occurs over iterative steps rewarding values produced on trials in which the agents' values match, and this results in convergence behaviors. The speech experiment behavioral data reported here exhibits the same pattern of results. As shown in Fig. 4, when two dyad members start interacting with each other, at least one dyad member, if not both, shows shifts in their mean values in production parameters to become more similar to the other member they are conversing with, as compared to when they are speaking solo prior to interaction.

This is further confirmed by the similarities in distribution patterns between the model simulation results (Fig. 1a & 1b) and the experimental data (Fig. 6). While convergence, measured as the decreased distance between the phonetic variables produced by each dyad member, is observed for all six behavioral cases reported here, only those five cases that have overlapping baseline values show a near-complete matching of the inter-partner values. In the other case, in which speakers have baseline values that are proximal but not overlapping, the speakers' production distributions still converge (and become overlapping) but remain distinct. This mirrors the cases of non-convergence in the model simulations, in which the distance between agents (with equal variability) is proximal and becomes closer by the time of the final iteration. This suggests that the natural range of individuals' production may constrain the extent of accommodation behaviors (Babel, 2010; Walker and Campbell-Kibler, 2015).

Both simulation and behavioral results support the hypothesis that speaker variability plays a key role in convergence. Three pieces of evidence are relevant here: speakers' baseline variability, the reduction in variability in speakers' behavior during interaction, and the relationship between the relative magnitude of shifts during convergence contributed by an individual speaker and that speakers' variability structure.

Our simulation results demonstrate that the intrinsically more variable agent of the two, i.e., the agent with a higher baseline variability, will be the 'converger.' The 70% of converged simulations in which the two agents have different baseline variabilities clearly captures this—the more variable agent always drives the convergence, showing a greater shift in mean than their partner (Fig. 1a-c). Given the larger window of variation in their production, the more variable agent can match the distant productions of their partner, thereby pushing their distribution towards that of their partner. Our experimental data (summarized in Table 2, baseline CoV column) also indicate this. In two of three dyads (Dyads S1-S2 & S3-S4), the more flexible dyad member has higher variability in solo speech before the interactive task than the less malleable dyad member.[3]

The remaining 30% of converged simulations are the cases in which both agents have identical values of baseline variability. Here, the relative contribution of each agent to convergence varies considerably across simulations—i.e., there is no preference for which agent shows a greater shift in mean during interaction (Fig. 2). These include cases such as only one agent shifting a substantial amount, the two agents shifting nearly equally, and all other intermediate possibilities. Our last converging dyad (S7-S8) from the experimental study appears to be an instance of this scenario; the two interlocuters with similar baseline variability values (baseline CoV column in Table 2) converge via nearly equally shifting their baseline mean values. As shown in Fig. 4E, F, while the shift in mean is slightly greater for Speaker S7 than for Speaker S8, both speakers converge towards each other in both measures. This is clearly so for sentence duration (Fig. 4E), whereas for the tongue tip stiffness measure (Fig. 4F), the dyad demonstrates a late match—i.e., convergence emerges in the later trials.

Our model further showed that the agent with higher baseline variability exhibits a substantially decreased variability during interaction (Fig. 1). This was also seen in four of six cases from our experimental results (Fig. 7). The more flexible speakers (S2, S4 & S7) typically show reduced production variability in the interaction task compared to speaking solo. That said, the two exceptions to this pattern (Fig. 7B, E) show fluctuations in both directions from the flexible speakers' baselines, exhibiting both increased and decreased variability. Contrary to our modeling results, which show stabilized production after the convergence process, these two exceptional experimental results suggest that the converger's speech may adapt to exhibit higher probabilities of producing a new central value without a reduction in variability. One possibility is that the more malleable speaker may be exploring various production targets within the wider parameter window available to them without reducing the size of that window.

Overall, the model predicts that the change in variability is greater for the more flexible agent who converges to resemble the distribution of the less flexible agent. This was confirmed in simulation results (Fig. 3) and can also be seen in our behavioral data. For Dyads S1-S2 and S3-S4, changes in variability from individuals' solo speech to interactive speech are consistently greater for the 'convergers' than for their less malleable dyad partners (Table 2, $|\Delta CoV|$ column), even though in one case (Dyad

[2] Variation in social and task-related factors can contribute to an asymmetry in accommodation (e.g., Abel and Babel 2017; Pardo et al., 2017; Taminga et al., 2016), but they were not the main focus of our study. In addition, the maze navigation task we employed here assigned both dyad members giver and receiver roles in information exchange.

[3] One measure from Dyad S1-S2 (tongue tip stiffness) does not reach significance though still in the direction of the difference is as predicted by the model (less malleable S1 < more malleable S2). The failure to achieve significance can probably be attributed to the nature of the "distribution free" sign test, which measures the direction of effect on paired values, rather than directly assessing their numerical magnitude, combined with a small sample size yielding less power.

S1-S2's tongue tip stiffness measure) the change involves an increase in variability, not a decrease. The members of Dyad S7-S8, who have comparable baseline variability and thus contribute to convergence nearly equally, also show a comparable shift in variability during interaction. In sum, in both simulation results and these experimental findings, structured variability serves as an illuminating index of individual adaptability in convergence behavior.

Certainly many behavioral accommodation cases do exhibit patterns beyond those that have been simulated and observed in the current study; thus two limitations of this work must be noted. First, we underline that our simple conceptual model of two interacting interlocutors is specifically applicable to accommodation situations that exhibit convergence behaviors between interlocutors, as the simulations model how production values from two agents along a continuum could converge. Our model successfully simulates the convergence case of complete matching of interlocutor production values, as well as approximation examples in which some approach towards convergence still occurs but no perfect matching of prototypical values between interlocutors is observed. However, the current modeling yields convergence only in cases in which baseline distributions of the interlocutors are proximal to one another, which is what is seen in the behavioral data as well.

Our relatively simple model makes the strong prediction that the extent of convergence in any dyadic interaction should be based solely on variability structure of the speakers. However, such a perfect prediction is of course unlikely. First, accommodation patterns can vary across measures because individuals may attend to different speech characteristics in their partner to a greater or lesser extent and may have different natural proclivities in their own variation patterns. We thus presume that the presence, absence or degree of convergence will vary across tasks and measures. Second, our modeling assumptions and simulations are staged for interactional contexts in which social and other factors are playing a limited role. In various social contexts with different attitudes or statuses of interlocutors, while we would expect individual variability to still contribute to accommodation behaviors, we would not speculate as to the extent of such contributions in light of other influences of social and motivational factors. It would certainly be fruitful to examine more datasets that provide a wider range of talker variation, incorporating various conversational contexts and topics. Nonetheless, the limited results of our attunement simulations and speech production behavioral data suggest that intrinsic variability may play a significant role in determining whether or not phonetic convergence occurs.

5. Conclusions

Convergence in dyadic interaction demonstrates the real-time adaptability of speech behaviors. This study provides novel evidence that one key to understanding the cognitive basis for this adaptability can be found in the intrinsic variability exhibited by a speaker in the production of a phonetic unit. Based on several hypothesized cognitive components affording real-time accommodation, including individuals' variability, we construct a simplified computational model of attunement, abstracting away from other "social" factors that could influence the convergence process. The model succeeds in simulating the convergence behaviors of two conversing interlocutors ("agents") over time. These simulation results show that i) the intrinsically more variable agent of the two is the converger, ii) this converger, but not the other agent, shows a substantial decrease in variability during the convergence process, and iii) the converger shows a greater change in their variability structure than does the converger's partner. Analysis of parallel behavioral data from a conversational experiment mirrors the findings of the model simulations. As such, our findings demonstrate the important contribution of individual variability/flexibility to speaker adaptability and identify this structured variability as a factor in determining who converges in a spoken language interaction exhibiting accommodation .

CRediT authorship contribution statement

Yoonjeong Lee: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration. **Louis Goldstein:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - original draft, Writing - review & editing. **Benjamin Parrell:** Formal analysis, Visualization, Writing - review & editing. **Dani Byrd:** Conceptualization, Methodology, Funding acquisition, Investigation, Writing - review & editing, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Dr. Sungbok Lee and Dr. Samantha Gordon Danner for their help with the original data collection and analysis reported in Y, Lee et al. (2018). We thank Sarah Harper for helpful discussion, and the editor and two anonymous reviewers for helpful suggestions. The work was supported by NIH DC003172 (Byrd). NIH DC01797 and NSF IIS-1704167 supported YL during manuscript preparation and revision. The data relevant to this paper is publicly available via a Zenodo repository with a DOI of 10.5281/zenodo.1119284, and is available via the URL https://doi.org/10.5281/zenodo.1119284.

References

Abel, J., Babel, M., 2017. Cognitive load reduces perceived linguistic convergence between dyads. Lang Speech 60 (3), 479–502.
Abney, D.H., Paxton, A., Dale, R., Kello, C.T., 2014. Complexity matching in dyadic conversation. J. Exper. Psychol. 143 (6), 2304–2315.
Abney, D.H., Kello, C.T., Warlaumont, A.S., 2015. Production and convergence of multiscale clustering in speech. Ecol. Psychol. 27 (3), 222–235.
Babel, M., 2010. Dialect divergence and convergence in New Zealand English. Lang. Soc. 39 (4), 437–456.
Babel, M., 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. J Phon 40, 177–189.
Bourhis, R.Y., Giles, H, 1977. The language of intergroup distinctiveness. H. Giles (Ed.). Language, Ethnicity, and Intergroup Relations. Academic Press, London, pp. 119–135.
Byrd, D., 1996. A phase window framework for articulatory timing. Phonology 13 (2), 139–169.
Cohen Priva, U., Edelist, L., Gleason, E., 2017. Converging to the baseline: corpus evidence for convergence in speech rate to interlocutor's baseline. J. Acoust. Soc. Am. 141, 2989–2996.
Fowler, C.A., Magnuson, J.S., 2012. Speech perception. M. J. Spivey, K. McRae, & M. F. Joanisse Cambridge Handbooks in psychology. The Cambridge Handbook of Psycholinguistics. Cambridge University Press, Cambridge, pp. 3–25.
Giles, H., 1973. Accent mobility: a model and some data. Anthropol. Linguistics 15, 87–105.
Giles, H., 2008. Communication accommodation theory. L. A. Baxter, & D. O. Braithewaite Engaging Theories in Interpersonal communication: Multiple perspectives. Sage Publications, Inc, pp. 161–173.
Giles, H., Coupland, N., Coupland, J., 1991. Accommodation theory: communication, context, and consequence. H. Giles, J. Coupland, & N. Coupland Studies in Emotion and Social interaction. Contexts of accommodation: Developments in Applied Sociolinguistics. Cambridge University Press, Cambridge, pp. 1–68.
Goldinger, S.D., 1998. Echoes of echoes? An episodic theory of lexical access. Psychol. Rev. 105, 251–279.
Goldstein, L., 2003. Emergence of discrete gestures. Proc. Int. Cong. Phonetic Sci. 15, 85–88.
Goldstein, L., Fowler, C.A., 2003. Articulatory phonology: a phonology for public language use. N. Schiller and A. Meyer Phonetics and Phonology in Language Comprehension and Production. Mouton de Gruyter, Berlin, pp. 159–208.
Goldstein, L., Nam, H., Kulthreshtha, M., Root, L., Best, C., 2008. Distribution of tongue tip articulations in Hindi versus English and the acquisition of stop place categories. In: Presented at *Laboratory Phonology*. Wellington, New Zealand, 11.
Gordon Danner, S., Vilela Barbosa, A., Goldstein, L, 2018. Quantitative analysis of multimodal speech data. J. Phon. 71, 268–283.

Guenther, F.H., 1994. A neural network model of speech acquisition and motor equivalent speech production. Biol. Cybern. 72, 43–53.
Guenther, F.H., 1995. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. Psychol. Rev. 102 (3), 594–621.
Harper, S., 2020. Consistency in difference: the relationship between articulatory variability and segment differentiation for individual speakers. In: Presented at *Laboratory Phonology*. Vancouver, British Columbus, 17.
Johnson, K., 1997. Speech perception without speaker normalization: an exemplar model. K. Johnson & J. W. Mullennix Talker Variability in Speech Processing. Academic Press, San Diego, pp. 145–165.
Johnson, K., Ladefoged, P., Lindau, M., 1993. Individual differences in vowel production. J. Acoust. Soc. Am. 94 (2 Pt 1), 701–714.
Keating, P.A., 1990. The window model of coarticulation: articulatory evidence. J. Kingston & M. Beckman Papers in Laboratory Phonology I. Cambridge University Press, Cambridge, pp. 451–470.
Keating, P.A., 1996. The phonology-phonetics interface. U. Kleinhenz Interfaces in Phonology. Akademie Verlag, Berlin, pp. 262–278. Studia grammatica 41.
Lee, Y., Gordon Danner, S., Parrell, B., Lee, S., Goldstein, L., Byrd, D., 2018. Articulatory, acoustic, and prosodic accommodation in a cooperative maze navigation task. PLoS One 13 (8), e0201444.
Levitan, R., Hirschberg, J., 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In: *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, pp. 3081–3084.
Lewandowski, E.M., Nygaard, L.C., 2018. Vocal alignment to native and non-native speakers of English. J. Acoust. Soc. Am. 144, 620–633.
Liberman, A.M., Whalen, D.H., 2000. On the relation of speech to language. Trends Cogn. Sci. (Regul. Ed.) 4 (5), 187–196.
Meltzoff, A.N., Moore, M.K., 1977. Imitation of facial and manual gestures by human neonates. Science 198 (4312), 75–78.
Meltzoff, A.N., Moore, M.K., 1997. Explaining facial imitation: a theoretical model. Early Dev Parent 6 (3–4), 179–192.
Nielsen, K., 2011. Specificity and abstractness of VOT imitation. J Phon 39, 132–142.
Niziolek, C.A., Nagarajan, S.S., Houde, J.F., 2013. What does motor efference copy represent? Evidence from speech production. J. Neurosci. 33 (41), 16110–16116.
Oudeyer, P.-Y., 2006. Self-organization in the Evolution of Speech. Oxford University Press, Oxford.
Perkell, J.S., Lane, H., Ghosh, S., Matthies, M.L., Tiede, M., Guenther, F., Ménard, L., 2008. Mechanisms of vowel production: auditory goals and speaker acuity. In: Proceedings of the 8th International Seminar on Speech Production, pp. 29–32.
Pardo, J.S., 2006. On phonetic convergence during conversational interaction. J. Acoust. Soc. Am. 119 (4), 2382–2393.
Pardo, J.S., Gibbons, R., Suppes, A., Krauss, R.M., 2012. Phonetic convergence in college roommates. J Phon 40 (1), 190–197.
Pardo, J.S., Urmanche, A., Wilman, S., Wiener, J., 2017. Phonetic convergence across multiple measures and model talkers. *Attention*. Percept Psychophys 79 (2), 637–659.
Pickering, M.J., Garrod, S., 2004. Toward a mechanistic psychology of dialogue. Behav. Brain Sci. 27 (2), 169–190.
Pierrehumbert, J.B., 2001. Stochastic phonology. Glot Int. 5 (6), 195–207.
Roon, K.D., Gafos, A.I., 2016. Perceiving while producing: modeling the dynamics of phonological planning. J. Mem Lang. 89, 222–243.
Saltzman, E., Byrd, D., 2000. Task-dynamics of gestural timing: phase windows and multifrequency rhythms. Hum. Mov. Sci. 19, 499–526.
Taminga, M., MacKenzie, L., Embick, D., 2016. The dynamics of variation in individuals. Linguistic Variation 16, 300–336.
Tiede, M., Bundgaard-Nielsen, R., Kroos, C., Gibert, G., Attina, V., Kasisopa, B., Vatikiotis-Bateson, E., Best, C., 2010. Speech articulator movements recorded from facing talkers using two electromagnetic articulometer systems simultaneously. Proc. Meetings Acoustics 11, 060007-060016.
Vatikiotis-Bateson, E., Barbosa, A.V., Best, C.T., 2014. Articulatory coordination of two vocal tracts. J Phon 44, 167–181.
Villacorta, V.M., Perkell, J.S., Guenther, F.H., 2007. Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. J. Acoust. Soc. Am. 122 (4), 2306–2319.
Tiede, M., Mooshammer, C., 2013. Evidence for an articulatory component of phonetic convergence from dual electromagnetic articulometer observation of interacting talkers. Proc. Meetings Acoustics 19 (1), 060138–060144.
Walker, A., Campbell-Kibler, K., 2015. Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task. Front. Psychol. 6, 1–18.
Whalen, D.H., Chen, W.R., Tiede, M.K., Nam, H., 2018. Variability of articulator positions and formants across nine English vowels. J Phon 68, 1–14.