Discourse Processes, 00:1–19, 2015

ISSN: 0163-853X print/1532-6950 online
DOI: 10.1080/0163853X.2015.1023965

Routledge
Taylor & Francis Group

Speech Accommodation Without Priming: The Case of Pitch

Tom Gijssels
Department of Psychology
University of Chicago
Department of Linguistics
Free University of Brussels
Brussels, Belgium

Laura Staum Casasanto
Department of Linguistics
University of Chicago

Kyle Jasmin
Laboratory of Brain and Cognition
National Institute of Mental Health
Institute of Cognitive Neuroscience
University College London, London, UK

Peter Hagoort
Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands
Donders Institute for Brain, Cognition and Behaviour, Nijmegen, Netherlands

Daniel Casasanto
Department of Psychology
University of Chicago

Correspondence concerning this article should be addressed to Daniel Casasanto, Department of Psychology, University of Chicago, 5848 S. University Avenue, Chicago, IL 60637, USA. E-mail: casasanto@alum.mit.edu

People often accommodate to each other's speech by aligning their linguistic production with their partner's. According to an influential theory, the Interactive Alignment Model, alignment is the result of priming. When people perceive an utterance, the corresponding linguistic representations are primed and become easier to produce. Here we tested this theory by investigating whether pitch (F0) alignment shows two characteristic signatures of priming: dose dependence and persistence. In a virtual reality experiment, we manipulated the pitch of a virtual interlocutor's speech to find out (1) whether participants accommodated to the agent's F0, (2) whether the amount of accommodation increased with increasing exposure to the agent's speech, and (3) whether changes to participants' F0 persisted beyond the conversation. Participants accommodated to the virtual interlocutor, but accommodation did not increase in strength over the conversation and disappeared immediately after the conversation ended. Results argue against a priming-based account of F0 accommodation and indicate that an alternative mechanism is needed to explain alignment along continuous dimensions of language such as speech rate and pitch.

INTRODUCTION

Speakers often accommodate to each other's speech. In the laboratory and in natural conversation, speakers tend to align multiple dimensions of their linguistic behavior, including their choices of words (Barr & Keysar, 2002; Bortfeld & Brennan, 1997; Niederhoffer & Pennebaker, 2002), syntactic constructions (Gries, 2005), allophones (Alim, 2004; Pardo, 2006), speech rate (Finlayson, Lickley, & Corley, 2012), and pitch (Gregory & Webster, 1996). Why do people change their speech to be more like their interlocutors'?

Priming as the Mechanism of Alignment: The Interactive Alignment Model

According to one influential theory, the Interactive Alignment Model (IAM; Pickering & Garrod, 2004), speakers align their speech with one another because of a simple priming mechanism. When speakers perceive an utterance, the activation level of specific linguistic representations is boosted. Consequently, when it is the speaker's turn to respond, the heightened activation of these representations increases the likelihood they will be produced in the response. A key benefit of such a priming mechanism is that it lightens the computational burden on the speaker: Linguistic representations are already activated in the comprehension process, so it becomes easier for the speaker to produce utterances based on these same representations.

If this account is correct, then alignment[1] should show two characteristic signatures of priming (see Wiggs & Martin 1998). First, alignment should be "dose dependent": the more often a listener perceives a given linguistic structure in a conversation, the higher the likelihood of producing that structure (Garrod & Pickering, 2004). The IAM predicts that "[a]s the conversation proceeds, it will become increasingly common to use exactly the same set of computations," (Garrod & Pickering, 2004, p. 10). Repeated priming of a representation over the course of a conversation should incrementally heighten its activation, leading to incrementally increasing alignment (Hartsuiker, Kolk, & Huiskamp, 1999). Second, if priming is the mechanism that drives alignment, then alignment effects should persist beyond the local exposure context. That is, once the activation level of a representation has been heightened, this activation should not immediately return to its baseline level; rather, it should remain heightened for some measurable period of time after exposure to the priming stimulus ends.

Both of these predictions about priming have been borne out in studies on syntactic alignment. Generally, these studies show that when participants hear or read sentences that contain one of two grammatically allowed constructions (e.g., active vs. passive voice), they are more likely to produce the primed construction (e.g., Branigan, Pickering, & Cleland 2000). Support for the dose dependence of syntactic alignment comes from several experiments that measured or manipulated the frequency with which participants encountered one of two syntactic alternatives. For example, the number of passive sentences speakers produce can be predicted by the number of passive structures the speakers produced or perceived previously (Jaeger & Snider, 2008; see also Kaschak, Loney, & Borreggine 2006).

Support for the persistence of syntactic priming comes from observations of syntactic alignment effects after delays ranging from 15 minutes up to 7 days after the initial priming manipulation (Kaschak, Kutta, & Schatschneider, 2011; Kaschak, Kutta, & Coyle, 2014). Priming also persists across changes in location or experimental context (Kutta & Kaschak, 2012).

Can Priming Explain Alignment of Continuous Features of Language?

Although priming appears to account for alignment of word choices and syntactic structures (Pickering & Garrod 2004), the IAM also predicts alignment of other features of speech for which automatic priming mechanisms are not easy to specify. Consider, for instance, alignment along continuous dimensions like pitch

[1]We use "accommodation" as an umbrella term referring to speakers' adaptation to their interlocutors' speech patterns (Giles, Taylor, & Bourhis, 1973, p. 178). We use "alignment" to refer specifically to convergent accommodation.

(Gregory & Webster 1996) and speech rate (Giles, Coupland, & Coupland, 1991; Finlayson et al., 2012). The IAM suggests the proposed priming mechanism "*applies at all linguistic levels*" (Garrod & Pickering, 2004, p. 9 [italics added]), including phonetic and phonological levels. The authors specify that "interlocutors align accent and speech rate" and that "when all levels are aligned, interlocutors will repeat each others' expressions in the same way (e.g., with the same intonation)" (Pickering & Garrod, 2004, pp. 174–175).

Yet, there are reasons to doubt that alignment of pitch or speech rate is driven by priming. Because these features are continuous, aligning one's pitch or speech rate with an interlocutor's presumably does not involve activating representations of linguistic units that match the ones previously used, setting alignment of continuous variables apart from alignment of discrete units (e.g., words, syntactic structures). Moreover, alignment of continuous variables typically does not result in an exact *match* between speakers; A speaker's speech rate and F0 are adjusted incrementally in the direction of the interlocutor's, but this adjustment does not typically result in production of *the same* rate or pitch (e.g., Staum Casasanto, Jasmin, & Casasanto, 2010). Finally, to our knowledge, there is no evidence that aligning on a continuous dimension like pitch would make speech production *easier*, which is one of the main computational-level motivations of the priming-based alignment model.

Is Priming the Only Mechanism of Alignment?

According to the IAM, the "interactive alignment process is automatic and *only depends on simple priming mechanisms*" (Pickering & Garrod, 2004, p. 188 [italics added]). Although numerous studies of alignment for discrete linguistic units show standard priming effects (i.e., dose dependence or persistence) and therefore support the IAM, no study to date has tested for these signatures of priming in the alignment of continuous features of speech. Could alignment of continuous features require a different explanation than alignment of discrete linguistic units and require a mechanism other than the priming mechanism proposed by the IAM?

To find out, here we used an immersive virtual reality (VR) environment to test whether pitch alignment shows the two signatures of priming described above: dose dependence and persistence. Participants had a conversation with a virtual agent whose F0 was either digitally raised or lowered. We measured participants' F0 during the conversation and during pre- and post-experimental periods.

By using a virtual interlocutor whose speech only varied on the dimension of interest, we were able to test for alignment effects during spontaneous conversation while maintaining full experimental control and precisely manipulating F0, which would be impossible with a live human confederate. Previous studies have demonstrated alignment to virtual interlocutor for discrete units of language (i.e., syntactic structures; Branigan, Pickering, Pearson,

McLean, & Brown, 2011; Heyselaar, Hagoort, & Segaert, 2014), and our own previous VR study showed alignment along a continuous dimension (i.e., speech rate; Staum Casasanto et al., 2010) but did not test whether priming was the mechanism of the observed alignment.

We tested two hypotheses. According to the first hypothesis, speech alignment only depends on one mechanism: priming. This mechanism applies at all linguistic levels, from lexical and syntactic alignment to alignment along continuous dimensions like speech rate and pitch: an explicit claim of the IAM (Garrod & Pickering, 2004; Pickering & Garrod, 2004). We refer to this possibility as the Single Mechanism Hypothesis. If this hypothesis is correct, then different kinds of speech alignment should show similar signatures of priming. Like lexical and syntactic alignment, F0 alignment should increase over the course of conversation (i.e., it should show dose dependence) and should persist after the conversation ends.

Alternatively, priming may only explain alignment for some linguistic features: specifically, discrete features that become easier to produce as their representations become more activated. But priming may not explain alignment for other features for which alignment does not increase the ease of production. This second hypothesis predicts that, unlike syntactic or lexical alignment, F0 alignment should show neither dose dependence nor post-conversation persistence. We call this the Multiple Mechanisms Hypothesis.

If the Single Mechanism Hypothesis were supported, this result would validate a central claim of the IAM: that alignment at all linguistic levels depends on priming. If the Multiple Mechanisms Hypothesis were supported, however, this result would challenge the IAM because different mechanisms would be needed to explain alignment for different kinds of linguistic features (i.e., continuous vs. discrete features).

METHODS

Participants

Seventy-two members of the Radboud University community (24 male) participated in exchange for payment. Participants were all native speakers of Dutch between the ages of 16 and 30 and were tested in the immersive VR lab at the Max Planck Institute for Psycholinguistics.

Speech Stimuli

VIRTUO's speech was prerecorded by a male native Dutch speaker reading in a conversational tone from a script of statements and questions designed to

simulate a conversation about products in a grocery store. Because F0 varies with speaker sex, we constructed a separate set of stimuli for a female virtual agent, VIRTUA. A female native Dutch speaker reproduced the questions VIRTUO posed. After hearing the male recording of each question, she was instructed to say the same phrase in a natural pitch but matching the intonation and stress patterns of the original recording as closely as possible. The F0 of both the male and female recordings was manipulated without changing the speed of the originals, using the "change pitch" function in the software package Audacity (http://audacity.sourceforge.net/). Participants in the High condition heard these recordings with an F0 raised by 5%, and those in the Low condition heard them lowered by 5%.

Virtual Environment

The virtual environment (VE) was a supermarket, which was custom-designed for this experiment using Adobe 3ds Max 4 (Adobe Systems Inc., San Jose, CA) software. The virtual supermarket consisted of a single long aisle with shelves on both sides, stocked with products, providing a variety of items for VIRTUO to inquire about. The experiment was programmed and run using WorldViz's Vizard software (WorldViz LLC., Santa Barbara, CA). Participants wore an NVIS nVisor SX60 (NVIS Inc., Reston, VA) head-mounted display (HMD), which presented the VE at 1,280×1,024 resolution with a 60-degree monocular field of view. Mounted on the HMD was a set of eight reflective markers linked to a passive infrared DTrack 2 motion tracking system from ART Tracking (Advanced Realtime Tracking GmbH, Weilheim, Germany), the data from which was used to update participants' viewpoints as they moved their heads.

Sounds in the VE, including the voice of the avatar, were rendered with a 24-channel WorldViz Ambisonic Auralizer System. The sound system was supplemented by four floor shakers mounted on a raised platform. These produced vibrations that contributed to an illusion of motion as participants were driven through the supermarket by VIRTUO in a specially modified virtual golf cart. VIRTUO was represented by a stock male avatar produced by WorldViz. The male avatar appeared to be a white male in his mid-twenties (the average age guessed by participants in debriefing was 26 years), which matched the age of the Dutch speaker who recorded his speech. VIRTUA was represented by a stock female avatar produced by WorldViz of a white female in her mid-twenties (the average age guessed by participants in debriefing was 26 years), which also matched the age of the Dutch speaker who recorded her speech.

Procedure

Before entering the VE, participants were told they would be having a conversation with VIRTUO or VIRTUA, a virtual agent who wanted to learn

more about the human world. They entered the VE by putting on the HMD, which showed them a virtual supermarket. When participants moved their heads, the display changed so they could explore the virtual world by looking around. Participants remained seated on a chair throughout the experiment. They traveled through the virtual supermarket in a virtual golf cart with VIRTUO/A in the driver's seat, so there was no need for participants to walk to move down the aisle of the grocery store. Participants were randomly assigned to the High or Low speech condition automatically by the experiment program, so the experimenter was not aware of which condition participants would be in until the experiment had begun, to eliminate the possibility of experimenter expectancy effects influencing participants' F0 before they spoke with VIRTUO/A. Once the experiment began, all instructions were written so participants did not have any verbal interaction with the experimenter during the experiment.

The experiment consisted of a Pre-conversation block of turns followed by a Conversation block and a Post-conversation block. During the Pre-conversation turns, participants were alone in the VE, and had an opportunity to get accustomed to their surroundings. We collected a sample of speech during this time to use for Pre-conversation F0 measurement. To elicit speech, we gave participants written instructions (via the HMD) to look at four products on the shelves in front of them, one at a time, and describe each product briefly.

After the four Pre-conversation turns, participants met VIRTUO/A, who introduced him- or herself in a few sentences. VIRTUO/A then took participants on a tour of the grocery store, stopping at six items (bananas, ketchup, light bulbs, toothpaste, cat food, and beer) to ask them three or four questions about each one. The order of the items during the Conversation block was counterbalanced across participants, allowing us to analyze the effect of item order (how far into the experiment an item appeared) independent of item type (whether the participant was talking about bananas, beer, cat food, etc.). Participants responded with information about the identity of the products, what they were made of, how they are used in the human world, and so forth.

Participants' speech was recorded through a microphone suspended from the HMD. VIRTUO/A's speech behavior created a conversational setting, but s/he did not have the ability to understand or flexibly respond to participants' utterances. The experimenter listened to participants' responses from a control booth and pressed a button to advance VIRTUO/A to the next utterance in his or her script. VIRTUO/A's speech began after a random delay between 150 and 400 ms, so the experimenter's button-pressing (i.e., turn-taking behavior) could not directly influence the speech of the participant. If the next item in VIRTUO/A's script did not constitute a sensible response to something a participant said, the experimenter pressed a button that caused VIRTUO/A to say that s/he did not understand, and that they should move on.

At the end of the Conversation block, VIRTUO/A said goodbye to the participants. Immediately afterward, the Post-conversation block started in which a written prompt appeared on the screen thanking participants for their participation and asking them to describe the study for future participants.

Speech Analysis

The first turn of the Pre-conversation block was discarded for all participants to eliminate variation due to adjusting to the VE. All data were manually coded in the speech processing software Praat (Boersma & Weenink, 2011). For each participant, the responses to each of VIRTUO/A's recordings were marked in the original recording. Any disfluencies[2] that might affect F0 measurement were excluded (e.g., laughter, coughing, etc.). Then, we calculated mean F0 for each participant's utterances separately using the "Get Pitch" function in Praat. To accurately measure F0, we used separate F0 ranges for analyzing the male and female recordings based on the average F0 intervals for Dutch male and female speakers (male, 50–250 Hz; female, 80–350 Hz) (Aoju Chen, personal communication). The same pitch analysis was applied to each recording of VIRTUO's and VIRTUA's questions.

Statistical Analyses

All statistical analyses (unless otherwise noted) were performed using mixed-effects multiple regression models (Baayen, Davidson, & Bates, 2008) in R (R Development Core Team, 2011) and used the R packages lme4 (Bates, Maechler, & Bolker 2011) and languageR (Baayen, 2011). Reported probability values were estimated using posterior distributions for model parameters obtained by Markov Chain Monte Carlo (MCMC) sampling. The full details of these analyses, including parameter estimates and the results for the nonsignificant factors, are provided in Tables 1 through 8.

The main mixed-effects models of the results included the following fixed effects: Condition: High, Low (i.e., whether participants spoke with VIRTUO/A whose pitch was adjusted up or down); Conversation Block: Pre-Conversation, Conversation, Post-Conversation; and Gender: Male, Female. Random effects were included where appropriate and consisted of random intercepts for subject and for question. The dependent variable for each of these models was participants' F0, that is, the mean F0 of the participants' responses to each of VIRTUO/A's questions during the conversation. To control for the variation in

[2]All responses to one of VIRTUO's turns were also excluded because this turn was a joke, the responses to which were nearly all disfluent in some way.

the duration of each of the participant's responses, we included duration weights for each data point to the model. To construct these weights, we first calculated the total speech time per conversation block for each participant, effectively summing the durations of all individual data points. Then, we divided the duration of each individual data point by that conversation block's total duration. As such, the weights reflect the duration of each single response as a percentage of that participant's total time spent speaking.

We ran additional mixed-effects models to test for turn-by-turn alignment in each condition during Conversation. Fixed effects for this model were as follows: VIRTUO/A's F0: the values for VIRTUO/A's F0 for each of the questions during Conversation (these values were mean-centered separately for VIRTUA and VIRTUO, so both genders could be analyzed with the same model); Item Number: the order of the specific item VIRTUO/A talked about during the conversation (1 through 6, mean-centered); and Gender (Male, Female). Random intercepts were included for subject and item type (i.e., whether the conversation item was cat food, toothpaste, etc.; Question could not be used in this model because it corresponds completely to a trial's unique VIRTUO/A's F0 value). The dependent variable was the same as in the preceding model, namely participants' F0.

In all models we deviation coded the categorical fixed effects, which are reported here for interpretation of the unstandardized beta's: Condition (Low = −.5; High = +.5), Gender (Male = −.5; Female = +.5). In pairwise comparisons of conversation blocks the earlier block was always coded as −.5 and the later block as +.5.

RESULTS

Do Participants Align F0 to VIRTUO/A?

To see whether our manipulation affected participants' F0, we tested the interaction between Condition and Conversation Block. As this model indicated that the effect of Condition varied depending on the Conversation Block (interaction: $\chi^2(2) = 14.88$, $p = .0006$), we ran follow-up analyses performing pairwise comparisons between each combination of the conversation blocks.

To test whether participants aligned to VIRTUO/A, we compared participants' F0 in Pre-conversation versus Conversation. Whereas in the Pre-conversation block participants' F0 did not differ between condition (High: mean = 179.3 Hz, SEM = 3.6 Hz; Low: mean = 178.3 Hz, SEM = 4.0 Hz; no main effect of Condition: $\beta = 1.06$, pMCMC = .70; Table 1), during Conversation participants aligned their F0. As predicted, participants in the

TABLE 1
Output of MCMC Simulation Test of the Mixed-Effect Model of Condition Predicting Participants' F0 for Pre-Conversation

Factor	*Estimate*	*MCMCmean*	*HPD95lower*	*HPD95upper*	*pMCMC*
(Intercept)	164.63	164.30	156.70	171.25	.0001
Condition	1.057	3.87	−4.11	6.22	.70
Gender	86.01	86.02	80.53	91.61	.0001

High condition had an F0 that was on average 3.89 Hz higher than the F0 of participants in the Low condition (main effect of Condition: $\beta = 3.89$, pMCMC = .04; High: mean = 179.2 Hz, SEM = 1.49 Hz; Low: mean = 175.4 Hz, SEM = 1.58 Hz; Figure 1 and Table 2). This effect was qualified by a significant interaction between Condition (High; Low) and Conversation Block (Pre-conversation; Conversation) ($\beta = 2.77$, pMCMC = .006) and the absence of a main effect of Condition ($\beta = 2.42$, pMCMC = .14; Table 3).

We then investigated whether alignment happened on a *turn-by-turn* basis throughout the conversation by testing whether VIRTUO/A's F0 for each question predicted participants' F0 in their responses. Because

TABLE 2
Output of MCMC Simulation Test of the Mixed-Effect Model of Condition and Item Number Predicting Participants' F0 for Conversation

Factor	*Estimate*	*MCMCmean*	*HPD95lower*	*HPD95upper*	*pMCMC*
(Intercept)	165.15	165.26	114.47	219.71	.0002
Condition	3.89	3.87	0.32	7.53	.04
Item no.	−.99	−.99	−1.28	−.70	.0001
Gender	85.4	85.40	81.55	89.33	.0001
Condition: item no.	0.21	0.12	−.41	0.76	.56

TABLE 3
Output of MCMC Simulation Test of the Mixed-Effect Model of Condition and Conversation Block Predicting Participants' F0 for Pre-Conversation vs. Conversation

Factor	*Estimate*	*MCMCmean*	*HPD95lower*	*HPD95upper*	*pMCMC*
(Intercept)	163.78	163.78	162.09	165.45	.0001
Condition	2.42	2.41	−.94	5.46	.14
Conversation block	−1.50	−1.51	−2.49	−.46	.0052
Gender	85.51	85.51	82.16	88.88	.0001
Condition: conversation block	2.77	2.77	0.72	4.81	.006

TABLE 4
Output of MCMC Simulation Test of the Mixed-Effect Model of VIRTUO/A's F0 and Item Number Predicting Participants' F0 in the High Condition

Factor	*Estimate*	*MCMCmean*	*HPD95lower*	*HPD95upper*	*pMCMC*
(Intercept)	168.65	168.65	163.47	173.94	.0001
VIRTUO/A F0	.50	.50	.36	.64	.0001
Item no.	−.57	−.58	−1.14	−.025	.04
Gender	84.45	84.39	76.47	92.76	.0001
VIRTUO/A F0: item no.	.014	.017	−.05	.09	.63

TABLE 5
Output of MCMC Simulation Test of the Mixed-Effect Model of VIRTUO/A's F0 and Item Number Predicting Participants' F0 in the Low Condition

Factor	*Estimate*	*MCMCmean*	*HPD95lower*	*HPD95upper*	*pMCMC*
(Intercept)	163.47	163.46	158.16	168.55	.0001
VIRTUO/A F0	.60	.61	.43	.78	.0001
Item no.	−1.32	−1.32	−1.95	−.67	.0001
Gender	90.45	90.49	81.94	98.98	.0001
VIRTUO/A F0: item no.	.0084	.0077	−.079	.090	.86

VIRTUO/A's F0 also contains the variance because of the difference between the High and Low conditions, we constructed separate models for the data from each of these conditions. Both models showed a significant effect of VIRTUO/A's F0, indicating local alignment in this dimension, over and above the main effect of Condition (High $\beta = .50$, pMCMC $= .0001$; Low $\beta = .60$, pMCMC $= .0001$; Figures 2 and 3, Tables 4 and 5). For every change in Hz in VIRTUO/A's F0, participants F0 changed their F0 in the same direction by .5 Hz in the High condition and by .6 Hz in the Low condition.

TABLE 6
Output of the Linear Model Testing Condition Predicting Participants' F0 for Post-Conversation

Factor	*Estimate*	*SE*	*t* value	*Pr(>|t|)*
(Intercept)	158	2.44	64.72	.0001
Condition	−.10	4.60	−.022	.98
Gender	79.41	4.88	16.27	.0001

TABLE 7
Output of MCMC Simulation Test of the Mixed-Effect Model of Condition and Conversation Block Predicting Participants' F0 for Conversation vs. Post-Conversation

Factor	*Estimate*	*MCMCmean*	*HPD95lower*	*HPD95upper*	*pMCMC*
(Intercept)	160.54	160.52	159.03	162.04	.0001
Condition	1.85	1.87	−.94	4.79	.21
Conversation block	−6.05	−6.06	−7.02	−5.09	.0001
Gender	82.30	82.32	79.35	85.37	.0001
Condition: conversation block	−3.90	−3.90	−5.8	−1.89	.0002

TABLE 8
Output of MCMC Simulation Test of the Mixed-Effect Model of Condition and Conversation Block Predicting Participants' F0 for Pre-Conversation vs. Post-Conversation

Factor	*Estimate*	*MCMCmean*	*HPD95lower*	*HPD95upper*	*pMCMC*
(Intercept)	161.23	161.23	158.61	163.84	.0001
Condition	.47	.47	−4.48	5.56	.84
Conversation block	−7.55	−7.55	−10.12	−4.98	.0001
Gender	82.63	82.59	77.19	87.65	.0001
Condition: conversation block	−1.13	−1.10	−6.19	3.99	.66

Does F0 Alignment Show Characteristic Signatures of Priming?

First, we tested whether the alignment effect observed during participants' conversation with VIRTUO/A persisted into the Post-conversation block as predicted on a priming account. Whereas participants had aligned during Conversation, this effect had disappeared during the Post-conversation measurement (High; mean = 171.18 Hz, SEM = 6.93; Low: mean = 171.28 Hz, SEM = 7.28; no main effect of Condition for the model testing for Post-conversation; $\beta = -.10$, $t = -.022$,[3] $p = .98$; Figure 1, Table 6). This analysis was licensed by a significant Condition by Conversation Block (Conversation vs. Post-conversation) interaction ($\beta = -3.90$; pMCMC = .0002) and no main effect of Condition ($\beta = 1.85$, pMCMC = .21; Table 7). The disappearance of the alignment effect during Post-conversation is also evident from the lack of any Condition by Conversation Block interaction when comparing Pre-conversation to Post-conversation ($\beta = -1.13$, pMCMC = .66). The main effect of Condition in this model was not significant either ($\beta = .47$, pMCMC = .84; Table 8).

[3]We could not include random effects in the model testing for an effect of Condition in Post-conversation because there was only one data point per participant.

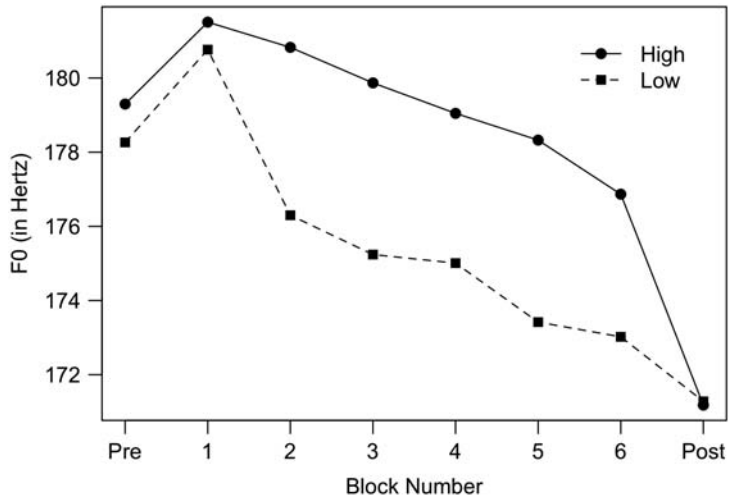

FIGURE 1 Participants' F0 by condition and conversation block.

Second, we tested whether the alignment effect showed dose dependence, that is, whether the strength of participants alignment increased over the course of the conversation. Whereas both Conversation models showed significant alignment effects as mentioned above, we found no evidence for any dose dependence in either model. The interaction between Condition and Item Number did not reach

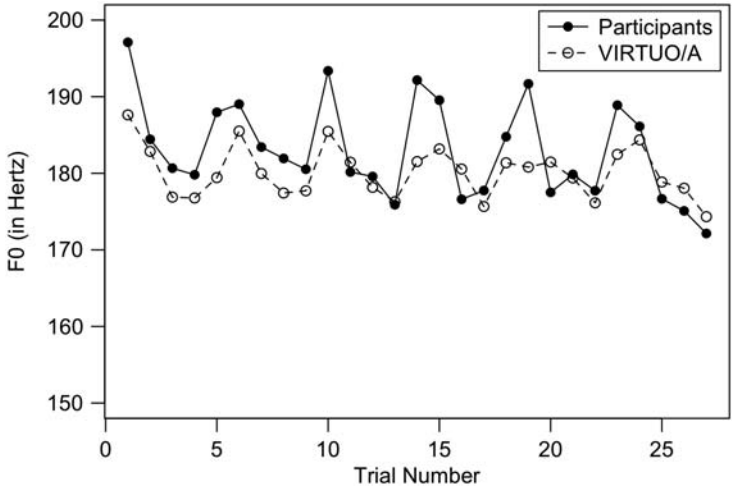

FIGURE 2 Participants' F0 and VIRTUO/A's F0 in the High condition by trial.

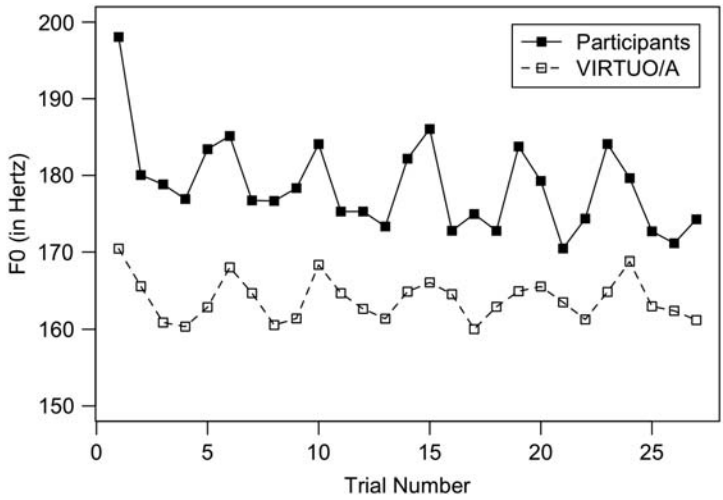

FIGURE 3 Participants' F0 and VIRTUO/A's F0 in the Low condition by trial.

significance ($\beta = .21$, pMCMC $= .56$; Figure 1), suggesting the effect of Condition did not vary over the course of the conversation. Similarly, in the turn-by-turn model, the interaction between VIRTUO/A's F0 and Item Number also did not reach' significance (Conversation High: $\beta = .014$, pMCMC $= .63$; Conversation Low: $\beta = .0084$, pMCMC $= .86$; Figures 2 and 3, Tables 4 and 5).

Additional Effects

Finally, in addition to the alignment effects of interest, several statistically significant patterns were expected but were not of interest with respect to our experimental hypotheses about dose dependence and persistence of alignment. First, all models showed an overall pitch declination effect: across all blocks, participants' F0 declined over time ($\chi^2(2) = 223.43$, $p = .0001$). This effect was also present in all pairwise comparisons: Participant F0 was, on average, 1.5 Hz lower in Conversation than during Pre-conversation ($\beta = -1.50$, pMCMC $= .0052$), 6.05 Hz lower in Post-conversation than in Conversation ($\beta = -6.05$, pMCMC $= .0001$), and 7.55 Hz lower in Post-conversation than in Pre-conversation ($\beta = -7.55$, pMCMC $= .0001$). Moreover, even within the Conversation block participants' F0 decreased: On average, with each additional item participants discussed, their F0 decreased by .57 Hz in the High condition and by 1.32 Hz in the Low condition (main effect of item number: Conversation High $\beta = -.57$; pMCMC $= .04$; Conversation Low: $\beta = -1.32$, pMCMC $= .0001$). Unsurprisingly, across all blocks we also observed a main

effect of Gender: females had a higher F0 than males ($\chi^2(1) = 192.8$, $p = .0001$; this effect is present in all pairwise models: Pre-conversation vs. Conversation: $\beta = 85.51$, pMCMC $= .0001$; Conversation vs. Post-conversation: $\beta = 82.30$, pMCMC $= .0001$; Pre-conversation vs. Post-conversation: $\beta = 82.63$, pMCMC $= .0001$). This shows that across all blocks, females had an F0 that was on average from 82.3 Hz to 85.51 Hz higher than male's F0.

DISCUSSION

Using immersive VR, we investigated whether speakers align their F0 to a virtual interlocutor and whether this effect can be accounted for by a simple priming mechanism. We found strong alignment effects: Participants who talked to a version of VIRTUO/A whose F0 was raised spoke with a higher F0, on average, than participants who talked to a VIRTUO/A whose F0 was lowered. Moreover, speakers matched their F0s to VIRTUO/A's dynamically on a turn-by-turn basis. However, the pattern of alignment showed neither of the two signatures of priming for which we tested: persistence and dose dependence. The alignment effect did not persist beyond the context of the conversation with VIRTUO/A; rather, it disappeared immediately once the conversation had ended. We also found no evidence that alignment was dose dependent: The alignment effect appeared immediately when the conversation started and did not increase (or decrease) in strength over the course of the conversation. As such, these data argue against priming as the mechanism of the F0 alignment effect.

These results support the Multiple Mechanisms Hypothesis and disconfirm the Single Mechanism Hypothesis, thus challenging a central claim of the IAM. Priming from perception of linguistic structures to production of the same structures may be the mechanism underlying *some* speech alignment effects (Pickering & Garrod, 2004) but not others. Specifically, we propose that priming should underlie alignment for discrete units of language like words and alternating syntactic constructions, but not for continuous features of language like speech rate and pitch.

Why should priming only underlie alignment for certain features of language? We suggest that one way to predict whether alignment is based on priming or an alternative mechanism is to ask whether accommodating along a given dimension is likely to make speech production *easier*. For dimensions like word selection or choice of syntactic alternation, the answer appears to be "yes," aligning with an interlocutor's choices should make speech production easier (i.e., less energetically costly, if not subjectively easier). By contrast, for dimensions like pitch and speech rate, the answer appears to be "no," aligning does not necessarily make speech production easier and could even make production more

difficult (e.g., imagine accommodating to an interlocutor whose speech is much higher or much faster than is comfortable for you).

How can the present findings be reconciled with previous studies showing persistent and dose-dependent alignment (e.g., Kaschak et al., 2006, 2014)? We do not believe our data invalidate these earlier findings, nor do they challenge the claim that these alignment effects for discrete features of language occurred because perceiving linguistic structures primed production of the same structures. Rather, we propose that not all linguistic alignment is produced by the same mechanism.

Different Kinds of Support for the Multiple Mechanisms Hypothesis

Although our data provide evidence against both signatures of priming for which we tested, the patterns of data that disconfirm persistence and dose dependence differ in their inferential power. The data argue strongly against persistent F0 alignment, which would be expected from a priming mechanism. The significant difference between conditions we found during the conversation was completely absent in the post-conversation phase, which immediately followed the conversation. This sets F0 alignment apart from other types of alignment, for which priming has been shown to persist for up to a week (Kaschak et al., 2014), and across physical contexts (Kutta & Kaschak, 2012).

By contrast, the evidence we present against a dose-dependent alignment is based on a non-difference. Alignment happened almost immediately at the start of the conversation, and we found no increase in the strength of alignment as the conversation progressed. This pattern replicates findings from a previous VR study in which we found that participants accommodated to VIRTUO's speech rate from the start of the conversation, without showing any increase in the strength of alignment over the course of the conversation (Staum Casasanto et al., 2010). The finding of robust but dose-invariant alignment to both F0 and speech rate suggests that the present data do not reflect a peculiarity of people's ability to accommodate along one particular dimension of speech but rather a generalizable pattern.

Functional Versus Mechanistic Explanations for Alignment of Continuous Features

If pitch alignment does not facilitate speech production, why do speakers accommodate to other speakers' F0s or to other continuous features of speech? According to Communication Accommodation Theory (Giles et al., 1991), alignment is often driven by the desire to communicate social goals and stances. Could this theory explain accommodation to a virtual interlocutor, who cannot understand these social moves? Perhaps. Although it is unlikely that our

participants thought their behavior could influence VIRTUO/A's beliefs about them, it is possible that the social nature of the conversation and the anthropomorphic agent led people to accommodate automatically. Some social behaviors seem to be so automatic they do not disappear in human–computer interaction even when they are totally illogical in these scenarios. For example, humans have been shown to exhibit politeness and reciprocity to computers (Fogg & Nass, 1997; Nass & Moon, 2000), leading Nass and colleagues to refer to these as *overlearned social behaviors*.

In this study, participants appear to have been enacting overlearned social behaviors, as evidenced by the disappearance of the alignment effect in the post-conversation block: from the moment the agent (and presumably the social motivation for alignment) was absent, people reverted to their baseline pitch. This immediate change is not compatible with priming (which decays gradually) but is compatible with sensitivity to changes in the social context. Alignment along at least some dimensions of linguistic behavior appears to be controlled by social factors that influence speakers' performance of continuous variables like pitch and speech rate rather than by the activation levels of discrete linguistic structures.

We note that by explaining alignment of continuous features of speech in terms of overlearned social behaviors but explaining alignment of discrete features in terms of priming, we are invoking different levels of explanation. Social motivations provide a *functional explanation* for alignment, whereas priming provides a *mechanistic explanation* (Bruce, 1985). These levels of explanations are not mutually exclusive: The alignment of both continuous and discrete features could be motivated by social factors, even though only the latter is subserved by a priming mechanism. So what is the mechanism underlying alignment of continuous features? We don't know. What we can conclude, on the basis of the present results, is that priming must not be the only mechanism of speech alignment: There must be different mechanisms of alignment for different kinds of linguistic features.

CONCLUSIONS

F0 alignment does not show dose dependence or persistence and is therefore unlikely to be the result of priming. These findings challenge a key claim of the most influential psychological model of speech alignment to date, the IAM (Pickering & Garrod, 2004), according to which priming is the sole mechanism of alignment. Priming appears to account for alignment of discrete linguistic structures, whose activation level rises cumulatively as they are used repeatedly, thus making their production easier. But priming does not account for alignment of the continuous dimension of language tested here: Hearing a particular F0 does not activate a discrete unit of language, and using the same F0 as one's interlocutor is not likely to make speech production easier. On the contrary, matching F0s could

be difficult (and also infelicitous) if one speaker's voice is naturally much higher or lower than their interlocutor's. Social motivations may provide a functional explanation for alignment along both discrete and continuous dimensions of speech, but a complete mechanistic understanding of speech accommodation will require additional mechanisms to be proposed and tested.

FUNDING

Supported by the Max Planck Gesellschaft, a James S. McDonnell Foundation Scholar Award (no. 220020236) to DC, and a PhD grant from the Scientific Research Fund of Flanders to TG.

REFERENCES

Alim, H. S. (2004). *You know my steez: An ethnographic and sociolinguistic study of styleshifting in a black American speech community*. Durham, NC: Duke University Press.

Baayen, R. H. (2011). languageR: Data sets and functions with "analyzing linguistic data: a practical introduction to statistics." R package version 1.4. Retrieved from http://CRAN.R-project.org/package=languageR

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

Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language*, *46*, 391–418.

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-42. Retrieved from http://CRAN.R-project.org/package=lme4

Boersma, P., & Weenink, D. (2011). Praat: Doing phonetics by computer [computer program]. Version 5.2.46. Retrieved 10 September 2011 from http://www.praat.org

Bortfeld, H., & Brennan, S. (1997). Use and acquisition of idiomatic expressions in referring by native and non-native speakers. *Discourse Processes*, *23*, 21–49.

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

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

Bruce, D. (1985). The how and why of ecological memory. *Journal of Experimental Psychology: General*, *114*, 78–90.

Finlayson, I., Lickley, R. J., & Corley, M. (2012, March). Convergence of speech rate: Interactive alignment beyond representation. In *Twenty-Fifth Annual CUNY Conference on Human Sentence Processing* (p. 24). New York, NY: CUNY Graduate School and University Center.

Fogg, B. J., & Nass, C. (1997). How users reciprocate to computers: An experiment that demonstrates behavior change. In *Proceedings of CHI 1997* (pp. 331–332). New York, NY: ACM Press.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*, 8–11.

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

Gregory, S. W., Jr., & 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*, 1231.

Hartsuiker, R. J., Kolk, H. H., & Huiskamp, P. (1999). Priming word order in sentence production. *Quarterly Journal of Experimental Psychology*, *52A*, 129–147.

Heyselaar, E., Hagoort, P., & Segaert, K. (2014). In dialogue with an avatar, syntax production is identical compared to dialogue with a human partner. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 2351–2356). Austin, TX: Cognitive Science Society.

Jaeger, T. F., & Snider, N. E. (2008). Implicit learning and syntactic persistence: Surprisal and cumulativity. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 1061–1066). Washington, DC: Cognitive Science Society.

Kaschak, M. P., Kutta, T. J., & Coyle, J. M. (2014). Long and short term cumulative structural priming effects. *Language, Cognition and Neuroscience*, *29*, 728–743.

Kaschak, M. P., Kutta, T. J., & Schatschneider, C. (2011). Long-term cumulative structural priming persists for (at least) one week. *Memory & Cognition*, *39*, 381–388.

Kaschak, M. P., Loney, R. A., & Borreggine, K. L. (2006). Recent experience affects the strength of structural priming. *Cognition*, *99*, B73–B82.

Kutta, T. J., & Kaschak, M. P. (2012). Changes in task-extrinsic context do not affect the persistence of long-term cumulative structural priming. *Acta Psychologica*, *141*, 408–414.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*, 81–103.

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, *21*, 337–360.

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

Pickering, M. J., & Garrod, S. (2004). Towards a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169–226.

R Development Core Team. (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Staum Casasanto, L., Jasmin, K., & Casasanto, D. (2010). Virtually accommodating: Speech rate accommodation to a virtual interlocutor. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 127–132). Austin, TX: Cognitive Science Society.

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*, *8*, 227–233.