Virtually accommodating: Speech rate accommodation to a virtual interlocutor.

Laura Staum Casasanto[1] (laura.casasanto@mpi.nl)
Kyle Jasmin[1] (kyle.jasmin@mpi.nl)
Daniel Casasanto[1,2] (daniel.casasanto@mpi.nl)

[1]Max Planck Institute for Psycholinguistics, Neurobiology of Language Group, Nijmegen, NL
[2]Donders Center for Brain, Cognition, and Behavior, Radboud University, Nijmegen, NL

Abstract

Why do people accommodate to each other's linguistic behavior? Studies of natural interactions (Giles, Taylor & Bourhis, 1973) suggest that speakers accommodate to achieve interactional goals, influencing what their interlocutor thinks or feels about them. But is this the only reason speakers accommodate? In real-world conversations, interactional motivations are ubiquitous, making it difficult to assess the extent to which they drive accommodation. Do speakers still accommodate even when interactional goals cannot be achieved, for instance, when their interlocutor cannot interpret their accommodation behavior? To find out, we asked participants to enter an immersive virtual reality (VR) environment and to converse with a virtual interlocutor. Participants accommodated to the speech rate of their virtual interlocutor even though he could not interpret their linguistic behavior, and thus accommodation could not possibly help them to achieve interactional goals. Results show that accommodation does not require explicit interactional goals, and suggest other social motivations for accommodation.

Keywords: Conversation; interaction; accommodation; alignment; virtual reality.

Introduction

Why do people accommodate to each other's linguistic behavior? Studies of multi-party interactions both in the laboratory and in natural conversation have suggested that two or more speakers in a conversation tend to align their linguistic behavior along several dimensions: lexical choice (Barr & Keysar, 2002; Bortfeld & Brennan, 1997; Niederhoffer & Pennebaker, 2002), phonetics (Alim, 2004; Pardo, 2006), and syntactic constructions (Gries, 2005), among others. The fact that accommodation occurs is well accepted, but the motivations for this convergence between speakers have been a matter of debate.

Studies of natural conversational interactions have identified social/interactional factors that influence how much speakers accommodate (Giles, Taylor & Bourhis, 1973). Based on these studies, it appears that a speaker accommodates towards or away from their interlocutor to achieve interactional goals: to make one's interlocutor do, think, or feel things. This can involve conveying social information, such as information about their social stances toward their interlocutor or toward a group that their interlocutor belongs to (Coupland, 1985). Accommodation could also help to coordinate joint actions being negotiated through conversation (Brennan & Clark, 1996).

But are immediate social motivations necessary to make speakers accommodate? Or might speakers accommodate even in the absence of a desire to achieve direct interactional goals? Mechanistic theories of dialogue (Pickering & Garrod, 2004) offer one possible alternative. Automatic alignment processes could account for convergence in linguistic behavior. That is, speakers might use similar linguistic forms to those used by their interlocutors because these forms are highly active and thus have an advantage over alternatives in the selection process.

On another alternative, accommodation could be a consequence of a speaker's attempt to achieve longer-term social goals: accommodation could be part of how speakers develop the linguistic styles that they use to communicate social information about themselves to others and to indicate their membership in social groups. This could occur in tandem with or independent of interaction-specific social goals.

Distinguishing these alternatives is difficult because in real-world conversations, interactional motivations are ubiquitous; in any conversation between two real people, the interlocutors may have social goals and relationships that could be influencing their linguistic behavior. Experimenters have attempted to deal with this complexity in a few ways. Experiments using pre-recorded speech in repetition paradigms have uncovered alignment between a speaker and a recording (Babel, 2009). However, because these experiments do not involve conversation, it's difficult to know whether the same mechanisms underlie speakers' production in these tasks and their accommodation in conversations.

In other experimental paradigms, the conversational setting is retained by using a confederate (Hannah & Murachver, 1999). However, no human confederate can entirely prevent his or her speech from being influenced by the naïve participant's speech. Introducing a confederate means losing experimental control over precisely those social and linguistic variables that might matter the most. This makes it difficult to assess the extent to which accommodation on the part of the participant depends upon their own interactional motivations, and to what extent it is a response to their interlocutor's behavior.

Virtual Reality (VR) provides an opportunity to engage participants in a conversational interaction with an interlocutor whose speech is not influenced by their speech, and can be varied systematically along a single dimension.

Moreover, a virtual interlocutor cannot feel or think at all, so participants cannot hope to influence the thoughts or feelings of the virtual interlocutor by accommodating to him. What happens in a conversational situation where interactional goals cannot be achieved? Do speakers still accommodate when their interlocutor cannot interpret their accommodation behavior? And if so, are they motivated by other, longer-term social goals, or is it a fully automatic process that is independent of social factors?

To find out whether speakers accommodate in a conversation with someone who cannot interpret their accommodation, we asked participants to enter an immersive VR environment and to converse with a virtual interlocutor, VIRTUO. While accommodation could theoretically occur along many dimensions at once, this experiment focused on the single dimension of speech rate, because this was easily manipulated in the virtual interlocutor. We varied VIRTUO's speech rate between participants to see whether participants would adjust their own speech rate to better match the rate at which their virtual conversational partner was speaking.

If immediate interactional goals motivate accommodation, then speakers in a conversation with VIRTUO should not accommodate to his speech rate, because they cannot hope to influence his thoughts, feelings or behavior by accommodating to him. If speakers *do* accommodate to VIRTUO by adjusting their speech rate towards his, there are two possible explanations: either accommodation occurs entirely automatically, or it can be motivated by social goals with a locus outside the current interaction (i.e., long-term social goals). To distinguish between these possibilities, we administered a post-experiment questionnaire investigating how participants judged VIRTUO on relevant social dimensions. If participants' judgments of VIRTUO correlate with their degree of accommodation to him, then this suggests that social goals beyond the level of the individual conversation influence accommodation.

Methods

Participants

Members of the Radboud University community (N=62, 30 male) participated in exchange for payment. Participants were all native speakers of Dutch between the ages of 17 and 28.

Materials

VIRTUO's speech was pre-recorded 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. The speed of the original recordings was manipulated without changing the pitch of the recordings using the "change speed" function in the audio manipulation software package Audacity, which removes or replicates short intervals of the acoustic signal in order to extend or shorten the overall length of a sound clip. Participants in the FAST condition heard these recordings sped up by 12%, and those in the SLOW condition heard them slowed down by 12%. Both sets of recordings remained within the range of possible speaking rates of a Dutch speaker, but the two rates were noticeably different.

The virtual environment (VE) was a supermarket, which was custom-designed for this experiment using Adobe 3ds Max 4. 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. Participants wore an NVIS nVisor SX60 head-mounted display (HMD), which presented the VE at 1280x1024 resolution with a 60 degree monocular field of view. Mounted on the HMD was a set of 8 reflective markers linked to a passive infrared DTrack 2 motion tracking system from ART Tracking, the data from which was used to update the participant's viewpoint as he moved his head. 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 4 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 avatar produced by WorldViz. The avatar's appearance suggested that he was a Caucasian male in his mid-twenties (the average age guessed by participants in debriefing was 26 years old), which matched the age of the Dutch speaker who recorded his speech.

Figure 1. VIRTUO in the virtual supermarket, from the perspective of a participant. The arrow indicates the next item that VIRTUO and the participants should discuss (here, ketchup). The steering wheel of the virtual golf cart is visible in the bottom left corner.

Procedure

Prior to entering the VE, participants were told that they would be having a conversation with VIRTUO, 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 in the drivers' seat, so there was no need for participants to walk in order to move down the aisle of the grocery store.

Participants were randomly assigned to the Fast or Slow speech condition automatically by the experiment program, so that the experimenter was not aware of which condition participants would be in until the experiment had begun. This minimized the possibility of experimenter expectancy effects influencing participants' speech rate before they spoke with VIRTUO. Once the experiment began, all instructions were written; therefore participants did not have any verbal interaction with the experimenter, which could have influenced their speech rate.

The experiment consisted of a Baseline block of trials followed by a Conversation block. During the Baseline trials, 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 as a Baseline speech rate. To elicit speech, we gave participants written instructions (via the HMD) to look at 4 of the products on the shelves in front of them, one at a time, and describe each product briefly.

After the four Baseline trials, participants met VIRTUO, who introduced himself in a few sentences. VIRTUO then took participants on a tour of the grocery store, stopping at six products (bananas, ketchup, light bulbs, toothpaste, cat food, and beer) to ask them three or four questions about each one. Participants responded with information about the identity of the products, what they were made of, how they are used in the human world, etc. Participants' speech was recorded through a microphone suspended from the HMD.

VIRTUO's speech behavior created a conversational setting, but 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 to the next utterance in his script. VIRTUO's speech began after a random delay between 150 and 400 ms, so that the experimenter's button-pressing (i.e., turn-taking behavior) could not directly influence the speech rate of the participant. If the next item in VIRTUO's script did not constitute a sensible response to something a participant said, the experimenter pressed a button that caused VIRTUO to say that he did not understand, and that they should move on.

Speech rate (in words per second) was calculated by transcribing participants' speech and marking the boundaries of their utterances as intervals in the audio and video transcription and coding software ELAN, then dividing the number of words transcribed by the number of seconds in the interval. Each participant's speech rate during the Conversation block was compared to their own Baseline rate.

Results and Discussion

Participants were assigned randomly to the two speed conditions, resulting in 33 participants in the Fast condition and 29 participants in the Slow condition. Mean speech rates during Baseline and Conversation periods are shown for participants assigned to the Fast and Slow conditions in Figure 2. Results indicate that VIRTUO's speech rate influenced how fast the participants spoke during their Conversation with him. Participants in the Fast condition spoke significantly faster during their Conversation with VIRTUO than during their Baseline measurement ($t(1,32)=4.02$, $p=.0003$), and significantly faster than participants in the SLOW condition ($t(1,60)=2.24$, $p=.03$), whose Conversational speech rate did not differ from their Baseline rate ($t<1$). This resulted in the predicted interaction between Condition (Fast, Slow) and Measurement (Baseline, Conversation; $F(1,60)=4.36$, $p=.04$; Figure 2).

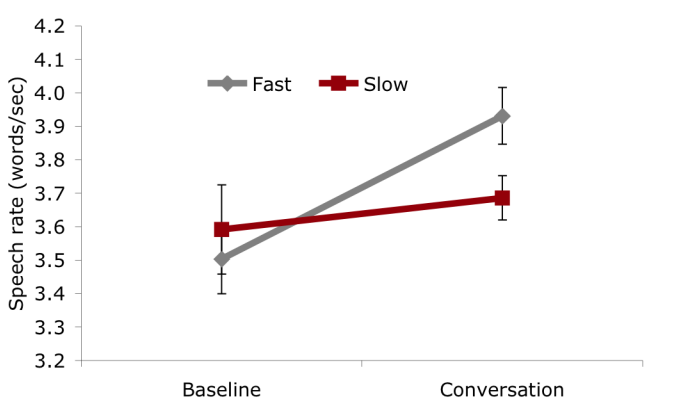

Figure 2. Experimental results. Speech rate differed between the Fast and Slow conditions during the Conversation period but not during the Baseline period.

The fact that the Baseline speech rate did not differ between conditions confirms that speaking with the experimenter prior to the experiment did not differently influence the speech of the Fast and Slow groups; rather, the differences that appeared in the Conversation period were a result of speaking to VIRTUO.

Participants in the SLOW condition did not speed up significantly, but there was a trend toward speaking faster in the Conversation condition than in the Baseline condition among these participants. This occurred despite the fact that their virtual interlocutor spoke slower than even their Baseline rate on average throughout their Conversation (VIRTUO spoke at 3.31 words per second on average in the Slow condition, and 4.20 words per second in the Fast condition). The slight increase in speed from Baseline to Conversation even among participants in the Slow condition suggests that while participants were influenced by VIRTUO's speech rate, they also may have been influenced

by other factors, such as increasing "immersion" in the virtual world (Heeter, 1992). This may have counteracted the effects of VIRTUO's speech rate on participants in the Slow condition. Importantly, the critical interaction between Condition and Measurement indicates that VIRTUO's speech rate affected participants' speech rate above and beyond any unexpected task-related effects.

To find out how early in the Conversation period accommodation could be detected, we first conducted an analysis comparing participants' speech rate in the Baseline period to their speech rate during their responses to VIRTUO's questions about the first item they discussed in their Conversation.

Results of this analysis suggest that speech rate can be adjusted quite quickly; a comparison of the Baseline speech rate with participants' speech in just the first item yields a significant interaction between Measurement and Condition ($F(1,58)=4.82$, $p=.03$), indicating that speakers in the Fast condition had already sped up more than speakers in the Slow condition over the course of the first 4 question-answer pairs. These results suggest that accommodation occurred rapidly and did not develop slowly over the course of the experiment.

This might seem surprising given the widespread assumption that accommodation is a process that occurs gradually over time. However, there are many respects in which speakers adjust to their interlocutors in the beginning of a conversation quite immediately; for example, when a Spanish-English bilingual speaker is approached by a stranger who begins to speak English to them, they are very likely to speak English in response immediately. There is no period of gradual adjustment in choice of language. Similarly, if someone begins a conversation with a friend in a sorrowful tone of voice, the friend is unlikely to respond back in a chipper tone; they will adjust to the emotional tone of the conversation immediately, without requiring a period of gradual change.

Participants' speech rate in response to VIRTUO's questions did not relate to the position in the experiment where the question appeared. This is consistent with rapid change immediately after meeting VIRTUO; perhaps speech rate accommodation does not occur gradually over time, but instead happens early in a conversation and is maintained fairly consistently throughout the interaction. However, the fact that accommodation did not increase over time in this experiment must be interpreted with caution, because the order of VIRTUO's questions was fixed rather than counterbalanced across subjects. Differences in content between the questions might have influenced participants' speech rate, which could have obscured any possible effects of the passing of time.

According to the questionnaire participants filled out after they finished the VR portion of the experiment, speakers accommodated more to VIRTUO when they judged themselves to be more similar to him ($r=.25$, $p=.05$). This correlation suggests that in addition to whatever automatic mechanisms might cause accommodation, people accommodate more to an interlocutor they identify with for longer-term social reasons (a point we will return to below).

General Discussion

The results of this experiment indicate that participants accommodated to the speech rate of their virtual interlocutor. Participants who spoke to a fast-talking VIRTUO sped up significantly from their Baseline speech rate, and spoke significantly faster than their counterparts in the Slow condition. This was true even though VIRTUO could not interpret participants' linguistic behavior, and thus accommodation could not possibly help them to influence VIRTUO's thoughts or behavior. Why, then, did participants accommodate to VIRTUO?

On one possibility, participants accommodated to VIRTUO through fully automatic mechanisms, without any social component. This would be consistent with studies showing alignment between a speaker and non-conversational speech (Babel, 2009; Goldinger, 1998), and might suggest that social motivations are unnecessary to cause accommodation in conversation.

However, results of the post-test support the idea that long-term social goals may be a factor that drives accommodation. Speakers may have social goals that extended beyond their current interaction. For example, some participants may have been motivated to accommodate to VIRTUO by a general tendency to speak similarly to other speakers, especially those that they can identify with to some extent.

Accommodation to certain interlocutors may be one of the mechanisms by which speakers develop a coherent linguistic style over a longer time scale, perhaps even playing a critical role in sound change (Niedzielski & Giles, 1996; Auer & Hinskens, 2005). Selective accommodation, to people with the right social characteristics, may help speakers speak in a way that reflects the way their in-group members speak.

The tendency to speak more like someone who one identifies with is fundamental to the organization of linguistic variation. Linguistic behavior at many levels, including phonetics, word choice, and choice of syntactic constructions, is subject to variation – there are many possible ways for a speaker to communicate approximately the same thing. Although this variation can seem random, these choices can often be predicted by a speaker's membership in various social groups. These groups can correspond to macrosociological categories (e.g. gender, age, ethnicity), or they can be locally defined (e.g. communities of practice) (Eckert & McConnell-Ginet, 1992). But how do these relationships between social group membership and linguistic behavior get established?

Individuals' linguistic behavior must be influenced by the behavior of others whom they consider to be in-group members in order for social groups to become correlated with linguistic behavior. The mechanisms underlying this process of sociolinguistic differentiation and identification are not entirely well-understood; however, they must

operate on the level of actual language use, i.e. individual conversations.

There is another way that social motivations could have influenced accommodation even in a virtual interaction. Perhaps people accommodated to VIRTUO because participants were somehow confused into thinking that he can interpret their social behaviors. VIRTUO does, after all, resemble a human interlocutor in many ways; perhaps speakers do not realize his limitations? It seems unlikely that participants were truly confused about this, given the restrictions in how VIRTUO could respond to them in this experiment. However, the principle that humans might interact with VIRTUO as though he were a real human even though he is not could still explain their accommodation behavior.

Some social behaviors seem to be so automatic that 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 (Nass & Moon, 2004), in what Nass and colleagues have called "overlearned social behaviors." If accommodation is such an overlearned social behavior, then people might accommodate to VIRTUO not because they think that they will influence his beliefs about them or behavior toward them, but because this behavior is applied automatically regardless of its applicability in a specific situation.

If this is in fact the reason why speakers accommodate to VIRTUO, then it suggests a reinterpretation of the accommodation we see in natural conversation as well. That is, interactional motivations may underlie linguistic accommodation, but in an automatic, overlearned way. If so, then speakers may not have specific intentions about the interaction they are engaged in, and they may have very little control over their accommodation behavior. This is consistent with the idea that accommodation can be motivated by general social goals, even in the absence of short-term social motivations.

Conclusions

In real-world conversations, accommodation may often be motivated by efforts to achieve interactional goals: people accommodate to make others do, think and feel things. But the present data show that this is not the only reason that people accommodate. Since people accommodate to a virtual interlocutor, we can conclude that accommodation is not necessarily driven by immediate attempts to influence social relationships or convey social messages. Yet, social motivations at a broader level may motivate accommodation, which may be a tool by which people develop linguistic styles, over the long term. The finding that the degree to which people accommodate correlates with how much they identify with their interlocutor suggests that accommodation is not merely a reflex. However, these results do not rule out some role for alignment processes that are engaged automatically. In real conversations, social and interactional factors may combine with automatic factors to produce linguistic accommodation.

Acknowledgments

Thanks to Albert Russel, Gerd Klaas, and Jeroen Derks for technical assistance in setting up the VR lab and programming the virtual world; to Sanne Berends, Daphne van Moerkerken, Merel van Rees Vellinga, and Tomas Bergvelt for help with data collection and coding; and to Pieter Seuren and Matthias Sjerps for help creating the verbal materials.

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 Univ. Press, 303 pgs.

Auer, P. & Hinskens, F. (2005). The role of interpersonal accommodation in a theory of language change. In P. Auer, F. Hinskens & P. Kerwswill (eds). *Dialect change: The convergence and divergence of dialects in European languages.* Cambridge University Press.

Babel, M. (2009). *Phonetic and Social Selectivity in Speech Accommodation.* Doctoral dissertation, Department of Linguistics, UC Berkeley.

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

Bortfeld, H. & Brennan, S. E. (1997). Use and acquisition of idiomatic expressions in referring by native and non-native speakers. *Discourse Processes*, 23, 119-147.

Brennan, S., & Clark, H. (1996). Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 22(6), 1482–1493.

Coupland, N. (1985). 'Hark, Hark, the Lark': Social motivations for phonological style-shifting. Language & Communication, 5(3):153-171.

Eckert, P. & McConnell-Ginet, S. (1992). Think practically and look locally: Language and gender as community-based practice. *Annual Review of Anthropology* 21:461–90.

Giles, Howard, Donald M. Taylor, and Richard Bourhis. (1973). Towards a theory of inter- personal accommodation through language: some Canadian data. *Language in Society* 2:177–192.

Goldinger, Stephen D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105:251–279.

Hannah, A. & Murachver, T. (1999). Gender and Conversational Style as Predictors of Conversational Behavior. *Journal of Language and Social Psychology,* 18(2):153-174.

Heeter, C. (1992). Being There: The Subjective Experience of Presence. *Presence: Teleoperators and Virtual Environments*, MIT Press.

Nass, C. & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues* 56(1):86-103.

Niederhoffer, K. and Pennebaker, J. (2002). Linguistic Style Matching in Social Interaction. *Journal of Language and Social Psychology*, 21(4).

Niedzielski, N., Giles, H., 1996. Linguistic accommodation (Essay no. 39). In: Goebl, H., Nelde, P.H., Stary, Z. and Wolck, W. (Eds.), *Contact Linguistics: An International Handbook of Contemporary Research.* De Gruyter, Berlin, pp. 33-342.

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

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