Casasanto, D., & Jasmin, K (2017). Virtual Reality. In A. de Groot & P. Hagoort (Eds.), Research Methods in Psycholinguistics and the Neurobiology of Language (pp. 174-189). Oxford: John Wiley & Sons.

Chapter 12: Virtual Reality

Daniel Casasanto and Kyle M. Jasmin

Abstract

Immersive virtual reality (iVR) is a rapidly developing technology through which experimenters can transport participants into virtual words. These worlds are rendered via stereoscopic video projections, which are typically enhanced with audio systems that simulate a 3-dimensional soundscape, haptic stimulators that make virtual objects seem tangible, and sometimes even olfactory stimulators. Traditional verbal or pictorial stimuli can induce experimental participants to imagine alternate realities; iVR can allow participants to experience them sensorially. Thus, iVR provides a degree of richness and realism that is not possible in traditional laboratory experiments, while enabling researchers to maintain rigorous control over the stimuli and the experimental environment. In this chapter we outline the basic components of iVR systems, discuss some ways in which they have been used to study social cognition, and describe ways in which this technology has begun to help researchers understand social aspects of language use.

Keywords: Virtual reality; Immersive; Psycholinguistics; Head-mounted display; Motion capture.

<A>Assumptions and Rationale

Language is the original virtual reality (VR) device. In the real world, what we can experience is limited by the richness of our surroundings, the reach of our arms, and the resolution of our senses. Through language, we can transcend these limitations and create an infinite number of alternate realities. Narratives can blast us into outer space (Asimov, 1951), plunge us 20,000 leagues under the sea (Verne, 1962), or lead us along a yellow-brick road toward an emerald-green city, past magic poppies and flying monkeys (Baum, 1958).

The worlds we create via language exist only in our imagination, and not in our senses. Information presented in other media, via newer kinds of "VR devices," can incrementally shift the burden of creating a virtual world from imagination to perception. Pictures in books and sound effects on the radio add unimodal (visual or auditory) details, both enhancing and constraining the imagined world. Audiovisuals on the stage, television, or in the movies supply even more perceptual details, yet the real world still exists alongside of the fictitious world. One need only glance away from the screen to return to reality, and remaining inside of these virtual worlds often requires a willing suspension of disbelief.

By contrast, in fully immersive virtual reality (iVR), which we describe below, the shift from imagination to perception is nearly complete. When people enter an iVR system the real world disappears, and an alternate reality commandeers the senses. What you see is determined by stereoscopic goggles that wrap around your field of view, and what you hear is determined by a montage of speakers that model a 3-dimensional soundscape. What you feel may be shaped by floorshakers beneath your feet, or vibratory feedback devices cued by your body movements. Some iVR systems even include olfactory stimulation.

How “immersive” are iVR systems? The answer depends in part on the system, and on the individuals’ propensity to feel “presence,” which is the term VR researchers use to describe one’s subjective immersion in the virtual world (Heeter, 1992). But a standard program that can run on even rudimentary iVR systems illustrates the grip iVR can have on most people’s minds. The “pit” illusion is simple. Participants stand at the mouth of a deep chasm, and are invited to walk across it on a plank of virtual wood. (Although it’s not necessary, some labs enhance the illusion by placing a real plank of wood on the ground at the participant’s feet – which lifts them about *one inch* above the floor.) The animation may not look realistic; the rocks and trees may look cartoony, and the 3D perspective may not be perfect. But still, the illusion may be inescapable. Many participants refuse to walk across the plank even though they *know* that there is absolutely no danger – that they are safely inside a university laboratory – and yet the mind cannot overrule the senses. There may be no need to suspend disbelief in iVR; disbelief may be impossible. (One of the authors of this chapter experienced severe vertigo the first time he crossed the plank, or rather failed to cross it.)

Aside from piquing people’s fear of heights, what is iVR good for? iVR offers a level of richness and realism that is difficult to achieve in the laboratory, while also letting researchers maintain rigorous experimental control over the stimuli and the experimental environment. Experimenters can stimulate multiple senses simultaneously, and collect multiple streams of data in parallel (e.g., vocal responses, body movement; also eye movement and electrophysiological data for iVR labs equipped with an eyetracker and electroencephalograph (EEG). By immersing participants in a virtual world, iVR may elicit more naturalistic responses to emotional or social stimuli than traditional methods do.

<A>Apparatus

The hardware supporting iVR can be divided into two types. *Input hardware* "captures" data from the real world, such as the position and motion of a subject's body. *Output hardware* "renders" the world to the subject by presenting some combination of visual, auditory and haptic information to the subject. In the middle, connecting the devices is a computer that processes the input and uses it to produce the output. We will take each type of device in turn.

<B>Input devices: motion-capture

Imagine you are seated in a virtual environment – a virtual classroom. You look at the person seated on your right, or perhaps, look down at your desk, where a virtual coffee mug is sitting. In doing so you of course move your head. Next, you pick up the coffee mug, and your virtual hand moves forward into your field of view, as it would in the real world. This is accomplished through the use of input technology called "motion capture" or "mo-cap."

Mo-cap allows the tracking of people and objects in the real world, for updating the positions of virtual people and objects in the virtual world. This is often done through the use of *markers*, small devices that attach to whatever body part or object one might wish to track. Two common types of markers – active and passive optical markers – rely on light and cameras to work. Passive markers are plastic balls with a reflective coating. They are called 'passive' because they do not themselves emit light; instead, they reflect light emitted from another source, such as an infrared lamp attached to the camera. Infrared is ideal for this purpose because it is invisible to the naked eye. Multiple cameras are used to pinpoint a marker's precise location and orientation in space.

Whereas passive markers reflect light, active markers emit it. Active marker systems typically consist of LED's worn on the body. As with passive markers, a camera detects the light and feeds this information to a computer in order to calculate the marker's location in space. With both types of systems, the more cameras you have, the better the results will be. This is true both because the triangulation of position can be more precise with more cameras, and also because markers only work when the camera can "see" them, that is, when they are not occluded or hidden. For example, suppose you are tracking the position of a subject's hand, and they reach behind their head. You would need a camera positioned to the rear of the subject in order for tracking to continue accurately.

A dataglove is capable of tracking movements of individual fingers. A classic but crude example is the Power Glove created by Nintendo in the 1980's. Professional datagloves used in virtual environments are more sophisticated, and are used for both input and output. Precise sensors in each finger of the glove allow a subject's hand shape and finger movements to be recorded. This data can be used to precisely measure hand gestures or linguistic signs and render the hand of an *avatar* (i.e., the character that embodies the participant in the virtual world) in real time. The glove can also serve as an output device by producing haptic feedback to simulate the sensation of holding or touching a virtual object. The dataglove does not transmit position information on its own, but by attaching a mo-cap marker to the glove, it is possible to locate the arm in the virtual environment.

A low-cost alternative to a full motion capture system is the Microsoft Kinect, which provides basic motion sensing. The system works without any markers at all; instead, a single camera positioned in front of the user detects motion against the background of the room, and infers both the user's position within the room and the position of their body. For some purposes, Kinect has been shown to work as well as more expensive optical systems (e.g., Chang et al., 2002).

You can also measure other kinds of behavior or physiology using equipment that is not specific to VR research. Microphones can be attached to the subjects to record their voice for later analysis (we will give an example of this below in Section 5).. Measures like eye tracking and galvanic skin response could also be incorporated.

<B>Output devices

Subjects are immersed in a virtual environment through output devices, which provide sensory information (visual, auditory, haptic) to the subject. Head-mounted Displays (HMDs) are a common method of presenting visual information. As the name implies, the device is worn on the head and consists of two video screens (one for each eye) attached to a helmet or visor. These screens project a first-person stereoscopic view that helps to create a 3-dimensional effect. The field of view varies. Generally, a device with a wider field of view allows more immersion and is more expensive. Some HMDs also provide head tracking through the use of accelerometers.

Although HMDs have in the past been expensive, low-cost options are emerging. Google released a product called "Google Cardboard," which was introduced in 2014 at the astonishing retail price of USD $15. It is a sheet of cardboard containing two lenses, which can be cleverly folded into a device that mounts a smartphone in front of the user's face (the smartphone is not included in the price). Together, the Cardboard and the smartphone make an effective HMD. The smartphone's screen is divided in two down the middle so that two images can be presented stereoscopically, one to each eye, to create a 3D effect. The phone's accelerometer provides head-tracking information so that the view of the virtual environment can be updated in real-time. A second low-cost device, the Oculus Rift, was released in 2016 at a price of USD $599. Rather than something you attach to your phone, the Rift is a full-fledged HMD. It provides a 110-degree field of view and built-in 3D headphones.

CAVE systems (short for *computer-activated virtual environments*) render virtual worlds without the need for an HMD. The environment is instead projected onto the walls, ceiling and floor of a room - similar to the “holodeck” from the Star Trek television series. The user wears 3D glasses that are synchronised with the projections on the sides of the CAVE, and separates the images into left and right for stereoscopy.

Presenting audio (e.g., voices) to subjects can be done with headphones built in to the HMD. Alternatively, external speakers can be placed on the walls, in the corners, on the flow in the ceiling, immersing the subject in a 3D sound experience. With this technique, the source location of sounds can be controlled exactly, if this is required.

<B>Moving through the virtual world.

How does a user move through a virtual world? The answer depends on the kind of physical constraints in your real-world laboratory, and the input and output hardware you use. If your laboratory is large enough, a subject can simply walk around the room (e.g., wearing an HMD and a backpack full of other hardware). Of course, any input and output devices the user may be wearing will need to stay connected to the computer, through either a wireless transceiver worn by the subject or through direct wired connections. Alternatively, wires can be fed straight up to a gantry system installed in the ceiling which moves around the room with the subject, keeping the right amount of slack in the wires. The position of the user in the real-world laboratory is tracked with motion capture (e.g., markers worn on the body), and this information is used to move the corresponding avatar in the virtual world.

Depending on the size of the VR lab, and whether the subject’s movement is, itself, of interest to the researchers, it might be better to let subjects sit still and move the environment around them. This option allows the virtual world to be infinitely large, even though the physical lab space is limited. In Staum Casasanto, Jasmin, and Casasanto (2010) and Gijssels et al. (2016), our subjects moved through a virtual supermarket. However, our lab was much smaller than a supermarket – in fact, participants could only take a few steps before reaching a wall. So instead of walking through the store, the avatar sat in a virtual motorized cart and was driven through the store by a virtual *agent* (i.e., an autonomous character in the virtual world – a digital robot). Floorshakers rumbled when the cart's virtual motor was operating, which provided haptic input and perturbed the subject's vestibular system to allow for an illusion of motion. Thus, the subject did not have to move through the lab – the virtual environment moved around them.

<B>Integrating input and output

Building your lab is the first step. The next is building your virtual world. Do you want your subjects indoors or outside? Do they need to walk around? Do they need to touch or manipulate objects? Will they talk to other people? The answers to these questions will affect your choices, but every virtual world needs one thing – a software system to integrate data from the input and output devices.

Although multiple software packages are available, one package popular among research psychologists is Vizard VR software, from WorldViz. It is an Integrated Development Environment (IDE) that controls multiple functions related to your experiment from within the same system or framework. With this tool, you can program what happens during your experiment and visually inspect the virtual world you are developing. During an experiment, the software handles program and data flow, processing input from motion capture cameras, microphones, and other streams, and updates the subjects' HMDs and audio headsets while they move their heads, hands, and bodies in the virtual world.

Vizard is based on the Python programming language, which may be advantageous to researchers who already use Python for other aspects of their research. In Vizard, virtual objects, avatars, and agents in the virtual world are all represented by Python "objects" that are easily controlled by changing their attributes (e.g., location=x,y,z; or color=blue) or activating their actions (making an agent "walk" or "speak", or a ball "drop"). When all of the various objects have been created for the world, controlling them with Python is only slightly more complex than other experiments such as video-game based tasks. Another benefit of Python is that it is open source, with many add-ons freely available.

The objects and avatars that populate your virtual world can be purchased or sometimes obtained free from a public repository. Software packages like Vizard sometimes come with a set of stock "models" (the specifications for the 3D object's physical shape) and "textures" (the bitmap graphics that map onto the model to give it its color and other visual attributes). Common situations, objects and people – for example, a man and a woman dressed in suits sitting at a conference in an office – will be easy to obtain. More niche needs (e.g., a pterodactyl flying past Macchu Picchu) will prove to be more difficult, and may require the aid of a graphic designer with experience working with 3D models.

<A>Nature of Stimuli and Data

In VR experiments, the virtual world itself is the stimulus, and it has nearly countless parameters to vary. You will need to choose which parameters to manipulate based on the exact experimental question or questions you are testing. Below, we will highlight some ways that aspects of virtual environments have been altered experimentally in the past and show how these paradigms could be adapted for language research.

<B>Manipulating parameters of virtual people

VR is effective when a person feels a strong "presence" in the virtual world, and responds to it as though it were real (Heeter, 1992). Establishing presence is what allows researchers to manipulate not just participants' sensory experience, but also their thoughts, beliefs, and behavior. VR allows us to change people's appearance in ways that are impossible in the real world. This can have consequences on a person's beliefs about themselves. A classic example is the "Proteus Effect". Yee and Bailenson (2007) altered the height of subjects' avatars. Some subjects were given a tall avatar, others a short one. They then played a competitive bargaining game. Subjects with taller avatars played aggressively, whereas those with shorter avatars were more likely to accept unfair deals. In another study, Fox et al. (2013) gave female participants either a conservatively-dressed avatar or one dressed in a revealing outfit. Participants who were assigned a sexualized avatar reported more body-related thoughts and reported more "victim-blaming" attitudes toward rape.

The Proteus Effect studies show that VR can be effective in altering people's beliefs about themselves. Could this effect be exploited for language research? If the height of a person's avatar activates stereotypes and affects their feelings of dominance and power, perhaps it could also affect their linguistic behavior as it relates to dominance. We might predict that people with taller avatars would behave more dominantly in conversation – talking louder, interrupting more, and accommodating less to the linguistic choices of the person they're speaking with. Conversely, a person with a shorter avatar might speak less loudly, interrupt less, and accommodate more to the language styles of their speaking partner. Changing an avatar's height is trivially easy in VR. Using Vizard software, you can simply specify in centimeters exactly how tall you would like a person to be.

There are other ways that changing how a subject appears might affect their linguistic output. Groom et al. (2009) showed that changing the race of an avatar can activate stereotypes and affect racial biases. Might changing the race of a participant also activate linguistic knowledge – words or phonological patterns associated with that race? Race could be varied simply by substituting one avatar for another. Manipulating the cultural subgroup of a subject through a change of virtual clothing could produce similar effects. (An aristocrat speaks differently from a hobo.) VR could prove to be a useful tool for exploring the extent of latent knowledge of other groups' linguistic patterns, and whether this knowledge can be activated and put into production by transiently changing a person's identity.

<B>Manipulating parameters of the environment

Perhaps you want a drastic change in the experimental environment: You can simply substitute one background environment for another. Previous studies have used this technique for effective mood manipulations. For example, Riva et al. (2007) created two park environments that were designed to elicit specific emotions. One featured inviting sounds, lighting, and textures designed to induce calm relaxation, while the other was darkly lit and used sounds and textures designed to evoke feelings of anxiety. These environments were effective at inducing the target moods. Indeed, the more presence the subject felt, the more this mood induction worked. Conversely, being immersed in one of these emotionally-charged parks also heightened feelings of presence (compared to being placed in a neutral park).

Why might it be useful to study language in different emotional contexts? There is some evidence that emotions affect language processing. Van Berkum et al. (2013) showed that moods induced with film clips (*Happy Feet* for a positive mood or *Sophie's Choice* for a negative one) affected the neural basis of pronoun reference assignment. VR could be used for more sophisticated mood inductions in the study of language processing, language production, and behavior in language interaction. VR allows greater experimental control than film clips, as the mood-inducing virtual scenes could be modified minimally to change the moods (in contrast to the use of film clips, which could differ along many different dimensions besides emotional valence).

VR mood inductions could also be useful for the creation of emotional vocal stimuli. Emotional vocal stimuli are often recorded by actors who merely pose the desired emotion, pretending to be fearful or relaxed, angry or excited. The actor is not actually experiencing the emotion they are trying to convey with their voice. This could be problematic if the portrayal is not convincing or if posed emotional vocalizations differ from real emotional vocalizations along some unknown dimensions. VR could be used to elicit genuinely emotional speech for an experiment. For the creation of fearful speech, experimenters could take advantage of the powerful “pit illusion” discussed in the introduction. People who experience a strong sense of presence in this illusion feel genuinely afraid. If they were asked to produce speech while they are experiencing the illusion, that speech should have all the characteristics of genuinely fearful speech.

Manipulating the spatial environment of a subject could also be useful for exploring relationships between language and space. Take for example reference frames for locating things in space. Languages like the Australian Guugu Yimitthir and Mexican Tzeltal use cardinal direction (north, south, east, west) to locate things in space, for example, “the ant is south of your leg” (Majid et al., 2004; Haviland et al., 1993). VR could be used to manipulate the physical environment to test how people keep track of their orientation with respect to the sun, geographic features like mountains, and so on, for the purposes of encoding spatial information in language.

Changing the visual background in an iVR experiment requires having more than one background and choosing which one to load for your experiment. The backgrounds can be designed in graphic editing and 3D-Modelling software.

<B>Nature of the data

What you decide to collect in terms of data is up to you and will depend on your experimental question. Just as you have myriad options for presenting and manipulating stimuli, the various input devices we discussed above allow much flexibility in data collection. If your experiment requires verbal responses, these will be picked up by the microphone and can be saved as WAV audio files (https://en.wikipedia.org/wiki/WAV) for linguistic or acoustic analysis. Any motion capture devices you employ will give you precise coordinates of where each marker was in space at each time point in your experiment. You can then time-lock these movements to events in your experiment, or other behavior (like vocalizations) and plot and analyze the movements.

<A>Collecting and Analyzing Data

As discussed above, using VR lets you have multiple data streams. You will have to decide what to collect and what to analyze. If your experiment uses motion capture, send position information for each of the markers to your log file, for the entire duration of your experiment. If you are recording audio from a microphone, record and save everything in a high quality uncompressed format. You may also want to record a video of everything your subject saw during the experiment. This is possible, but it will require a lot of disk space, so you will need to make sure you have a large hard drive with fast disk access.

Much of the data you collect can be analyzed using software you might already be familiar with. For example, if you are collecting audio recordings of subjects' voices, these can be analyzed with Praat (Boersma & Weenink, 2011), a well-established tool for measuring and manipulating aspects of voices. You could use Praat to, for example, measure pitch, inflection, and durational characteristics of subjects' voices. Movement-related information is recorded as millisecond-level timeseries of x, y, and z, coordinates for markers. You can compute quantities like velocity and acceleration in Matlab (Mathworks, Natick, MA). Alternatively, if only a simple analysis of movement is required for your experiment, such as where a subject gestured in left-right space, you could simply export movement data for the y-axis. This simple one-dimensional timeseries can be loaded into, for example, ELAN software (Brugman & Russel, 2004) and plotted with respect to other data streams such as audio and video recorded during the experiment and the timing of specific events.

<A>Exemplary studies

There is enormous potential for VR in language research, although there are relatively few published studies. We will highlight two examples and explain why using iVR was advantageous.

If we consider language to be a low-tech tool for creating virtual worlds, then non-immersive VR has been used to study language since the earliest experiments in psycholinguistics. Immersive VR, however, has been used in only a handful of psycholinguistic studies to date. A study by Gijssels, Staum Casasanto, Jasmin, Hagoort, and Casasanto (2016) tested the psychological mechanisms underlying linguistic accommodation (i.e., the tendency of speakers to adjust their linguistic production to be more (or less) like their interlocutor's; Giles, Taylor, & Bourhis, 1973). According to a leading psycholinguistic theory (Pickering & Garrod, 2004), all speech accommodation is the result of an automatic priming mechanism. According to this theory, called the Interactive Alignment Model (IAM), perceiving an utterance raises the activation level of the linguistic representations in the percept. Consequently, when it is the perceiver's turn to speak, the heightened activation of these representations increases the likelihood that these forms will be produced. Producing forms that have been primed by an interlocutor lightens the speaker's computational load; this is the functional motivation for accommodation, according to the IAM (Pickering & Garrod, 2004; see Chapter 6 for details about the priming methodology).

Gijssels and colleagues (2016) reasoned that, if priming is the mechanism of accommodation, then accommodation should show two “signatures” of priming: *dose dependence* and *persistence* (Wiggs & Martin, 1998). For alignment to be “dose dependent” means that the more often a listener perceives a given linguistic feature in a conversation, the higher the likelihood of producing that feature becomes (Garrod & Pickering, 2004). Thus, increasing exposure to a given aspect of linguistic production should cause accommodation to increase incrementally throughout a conversation (Hartsuiker, Kolk, & Huiskamp, 1999). For alignment to be “persistent” means that alignment effects should persist beyond the local exposure context. That is, once a feature of language has been primed, its heightened activation should not immediately return to its baseline level; rather, activation should remain heightened for some measurable period of time after exposure to the priming stimulus ends.

Both of these signatures of priming have been found in studies of syntactic accommodation: The more speakers were exposed to a construction (e.g., active vs. passive verb phrases) the more likely they were to produce the construction themselves (e.g., Branigan, Pickering, & Cleland 2000; Jaeger & Snider, 2008). Such syntactic alignment effects have been observed to last up to 7 days after the initial priming manipulation (e.g., Kaschak, Kutta, & Coyle, 2014), and to persist across changes in location or experimental context (Kutta & Kaschak, 2012). The IAM predicts that priming is responsible for accommodation effects “at all linguistic levels,” including continuous dimensions of language like *speech rate* and *pitch* (i.e., f0; Finlayson et al., 2012; Garrod & Pickering, 2004; Giles, Coupland, & Coupland, 1991; Staum Casasanto, Jasmin, & Casasanto, 2010). Because these features are continuous, aligning one’s pitch or speech rate with an interlocutor’s presumably does not involve activating representations of discrete linguistic units (e.g., words, syntactic structures) that match the units used previously by an interlocutor.

It seems unlikely, therefore, that priming is the mechanism of accommodation along continuous dimensions of linguistic production like speech rate and pitch, in which case accommodation effects should *not* show dose dependence or persistence. To test this prediction, Gijssels and colleagues (2016) measured the pitch of participants' speech before, during, and after their conversation with a virtual agent, in iVR. Male and female participants discussed items in a virtual supermarket with a lifelike virtual agent of their same gender (named VIRTUO or VIRTUA) at the iVR lab at the Max Planck Institute for Psycholinguistics, in Nijmegen, The Netherlands.

The supermarket environment was created specifically for this experiment using pre-made 3D models and textures that were integrated with Adobe 3ds Max 4 software (Adobe Systems Inc., San Jose, CA). We started with an empty supermarket model, then added shelves and products to put on the shelves. The VIRTUO and VIRTUOA characters were 'stock' models that came with Vizard Software.

The various items you typically find in a supermarket served as the topics of conversation. To make sure there were always new things to talk about, there needed to be new items in the immediate visible environment of the subject and the virtual conversation partner. This was accomplished by 'moving' the participant through the supermarket in a virtual vehicle. Subjects sat in a chair in the real world, which became a motorized golf cart in the virtual environment. VIRTUO/A sat behind the steering wheel and 'drove' the subject down the supermarket aisle. Floor shakers rumbled as the virtual engine ran, simulating the sound and feel of an engine.

Although this might seem quite complicated to set up, Vizard allows experimenters to control programming flow at a very high level. Moving a virtual golf cart can be as simple as specifying the golf cart's object ID and the coordinates it should move to (e.g., "golfcart.move([x, y, z] speed = s") and starting the engine ("floorshakers.Start"). The difficult part is setting up all of the hardware and software that makes this possible.

In the experiment, the agent asked the participant a series of questions about each item (e.g., *What is ketchup made of?*). VIRTUO's and VIRTUA's voices were recordings of native Dutch speakers of the same gender. Crucially, the F0 of these recordings was adjusted to be 5% higher or lower than the original, and participants were randomly assigned to interact with either the *high* or *low* version of VIRTUO/A. Pitch was manipulated with Audacity software, which is freely downloadable (http://audacity.sourceforge.net). An experimenter listened to the conversation between the participant and the agent, and triggered VIRTUO/A to make an appropriate response, at the appropriate time.

Results showed that, compared to a pre-experimental sample of speech (recorded while the participant was in the virtual world, but before they met VIRTUO/A), the pitch of participants' speech was adjusted in the predicted directions. Participants assigned to interact with the high VIRTUO/A spoke significantly higher, on average, than participants assigned to interact with the low VIRTUO/A. Moreover, the participants' F0s tracked the agents' F0s on a turn-by-turn basis. However, the magnitude of accommodation did not increase over the course of the conversation (i.e., with more exposure to the interlocutor's pitch), nor did it persist in the post-experiment sample of speech that was collected immediately after the conversation with VIRTUO/A ended. Thus, although participants showed a strong speech accommodation effect, accommodation showed neither dose dependence nor persistence, suggesting that priming was not the mechanism underlying this effect (see Staum Casasanto, et al., 2010, for a compatible finding in which participants accommodated their speech rate to match VIRTUO/A's). According to the IAM, speech alignment in all of its forms (e.g., lexical, syntactic, phonological) "is automatic and *only* depends on simple priming mechanisms" (Pickering & Garrod, 2004, p. 188, italics added). Yet, contra the IAM, Gijssels et al.'s (2016) results suggest that priming is not the only mechanism of speech accommodation, and that it is necessary to posit different mechanisms underlying different types of accommodation (i.e., accommodation along discrete vs. continuous dimensions of speech production).

Why did Gijssels and colleagues use iVR to address this question? First, it would be impossible to achieve the same level of experimental control with a human confederate, who could never modulate his or her F0 to be precisely 5% higher for half of the participants and 5% lower for the other half. Beyond pitch, it would be impossible to control myriad other physical and social aspects of the way confederates use their voices and their bodies, which could all potentially influence accommodation. All of these were held 100% constant across conditions with VIRTUO/A. Accommodation has been observed using a much simpler, non-immersive VR device, an audio recording (e.g., Babel, 2009), which allows for control of the voice but *eliminates* all other physical and social aspects of the conversation (e.g., gaze). Why not simplify this experiment and use an audio recording? Although an audio recording may be useful for answering some questions about conversation, language in its "natural habitat" is multimodal (not just auditory) and situated (interlocutors share a physical environment which constitutes an important component of their common ground; Clark, 1996). Stripping away the information that is typically available to language users as they *see* each other and their shared environment may blind researchers to important features of linguistic behavior. Accommodation exemplifies an aspect of language that is manifestly social (e.g., Babel, 2009; Giles et al., 1973), and may therefore be affected by extralinguistic aspects of an interaction. Accordingly, in an iVR study of speech-rate accommodation, Staum Casasanto et al. (2010) found that participants who rated themselves to be more similar to VIRTUO/A showed stronger accommodation effects.

As these experiments with VIRTUO/A illustrate, immersive VR can provide a rare combination of experimental control and richness or realism that is hard to achieve with human interlocutors or with simpler VR devices. But an important question remains open: Do the conclusions of experiments on conversation in iVR generalize to conversations between two humans? A study by Heyselaar, Hagoort, and Segaert (2015) addressed this question by testing whether using iVR to study syntactic accommodation yields similar results to studies using human speakers and listeners. They compared syntactic priming when humans were interacting with (i) other humans, (ii) humanlike virtual interlocutors, and (iii) computer-like virtual interlocutors.

Results showed that the rate at which participants produced passive vs. active syntactic constructions was affected equally by interacting with another human and by interacting with a humanlike agent. By contrast, this effect was reduced when the humans interacted with computer-like virtual interlocutors. These findings suggest that iVR with humanlike interlocutors presents the opportunity to study linguistic behavior with extraordinary experimental control over linguistic and extralinguistic aspects of the stimuli and the testing environment, without sacrificing the ability to generalize the results to real conversation between humans.

<A>Advantages and Disadvantages

Throughout this chapter we have emphasized that iVR allows for unprecedented levels of environmental richness and sensorimotor realism, while also enabling the experimenter to maintain strict control over myriad variables that would vary uncontrollably if human confederates were used rather than virtual agents or avatars. Here we mention some other potential advantages of iVR, as well as some disadvantages.

<B>Expanding the participant pool

Networked VR systems may allow greater diversity in the subject pool (Blascovich et al., 2002; Fox et al., 2009). As HMDs like the Oculus Rift become more affordable and commonplace, and with a fast internet connection, it should be possible to test participants remotely, without the typical geographic constraints imposed by the laboratory. Participants in different locations, perhaps with vastly different cultural or linguistic backgrounds, could interact within the same virtual environment.

Atypical populations would be one area of applicability. For example, people in residential care, who are unable to travel, would be able to put on an HMD and be transported anywhere, to talk to anyone, thus opening up possibilities for studying language processing and use in older people or people with mental disorders. A mobile VR lab is possible in principle, so long as motion capture needs are minimal, relying on, for example, an accelerometer in the HMD rather than external cameras to track head motion.

<B>Emotional realism

One of the challenges researchers face in studying emotion in the laboratory is that genuine emotions are difficult to elicit. Even strongly emotional words or pictures may fail to affect participants emotionally in the way real-life scenarios do. By commandeering the senses and immersing participants in virtual worlds, iVR may be useful for overcoming the emotional impotence of traditional stimuli. The pit illusion described before elicits real fear and anxiety. iVR may be capable of eliciting many other emotions as well. For example, even in non-immersive VR such as the Second Life online social environment (www.secondlife.com), interacting with other people's avatars can cause people to fall in love for real (Meadows, 2007).

<B>Reproducibility of complex environments

Much can vary between any two naturally occurring conversations, from the surroundings, background noise, weather, experimenter's clothes and behavior, and so on. iVR allows tight control over all sensory input delivered to the subject, such that the experience is replicated exactly for each subject (Blascovich et al., 2002; Fox et al., 2009). Verbal interactions between a person and a computer-driven agent can be structured and scripted such that the agent says exactly the same thing in each interaction, in exactly the same way, with all of the accompanying nonverbal behaviors held constant as well. In an interaction between two person-controlled avatars, the physical layout of the environment can be set up exactly the same for each experiment. Controlling the layout of objects in the environment could be especially useful for the study of reference (Keysar et al., 2000).

<B>Pitfalls of iVR

The realism of iVR can have its downsides. The illusion of height or of motion can be so powerful that it causes nausea, in a minority of subjects. Heyselaar et al.'s (2015) study (see above) raises another important consideration in iVR research: beware of creepy agents. People are somewhat comfortable interacting with robots that look nothing like humans (picture R2D2, the garbage-can–shaped robot in the *Star Wars* movies), and may be more comfortable interacting with anthropomorphic robots (like R2D2's tall golden sidekick, C3PO). But when robots or digital agents become *too* humanlike people typically have an aversive reaction: An anthropomorphic figure that succeeds in looking about 90% humanlike falls into the *uncanny valley* between the obviously artificial and the convincingly realistic (Mori, 1970). For example, humanlike prosthetic hands, which fall short of looking fully lifelike, are typically judged to be creepier than metal prostheses that are obviously not human. To ensure that their humanlike agent did not fall into the uncanny valley, Heyselaar et al. (2015) asked a group of raters to evaluate the candidate agents' faces, and they chose one that was rated high on humanness but low on creepiness. Stumbling into the uncanny valley could produce unexpected effects for any experiment with a social component.

Perhaps the greatest potential pitfall, if you are new to VR, is the investment of both time and money that can be required to create even a 'simple' iVR study. Although a portable HMD can be purchased cheaply (e.g., Google Cardboard), as can a simple motion tracking system (e.g., Microsoft Kinect), the virtual interactions you have in mind may or may not be feasible with a low-cost system. Detailed tracking of multiple body parts may require more sophisticated, multi-component mo-cap technologies. Even if you use stock characters as agents and avatars, creating the virtual world may require a substantial amount of programming, and populating it with 3-D models a substantial amount of artistry. Researchers new to iVR should be aware of the extent of equipment and expertise that may be needed to turn the study they are imaging into a (virtual) reality. On the other hand, the catalog of tasks that can be accomplished with low-cost hardware and pre-packaged software is growing quickly.

<A>Conclusions

Language researchers typically face a trade-off between experimental control and richness or realism of the experimental stimulus. Immersive VR can provide high levels of control *and* realism, compared to lower-tech methods of creating virtual worlds (e.g., words, pictures, video, and audio recordings). To date, iVR has been used in only a few psycholinguistic studies, to address questions about speech accommodation (as illustrated above) and gesture-speech interaction (Chu & Hagoort, 2014). Yet, in other areas of psychology iVR is already being used in imaginative ways, to address a variety of questions. Since language use is inherently interactive, iVR is a natural tool for language researchers to explore – one that allows experimental participants to interact with one or more interlocutors (other avatars or virtual agents) in a panoply of physical and social environments, while assuming diverse physical and social identities. Even if iVR environments or characters look somewhat artificial (thus avoiding the uncanny valley), they can elicit real emotions and social attitudes, allowing researchers to observe language in the kinds of socio-affective contexts in which it is typically used but rarely studied. With the advent of affordable motion capture and iVR technologies like the Microsoft Kinect, Google Cardboard, and Occulus Rift, mo-cap and iVR are no longer the province of those few researchers with access to a full-fledged VR laboratory. Like ERPs in the early 1980s and eye tracking in the late 1990s, iVR is now poised to become one of the psycholinguist's go-to methods.

Acknowledgments

We thank Laura Staum Casasanto for helpful discussions. This work was funded by a James S. McDonnell Foundation Scholar Award (#220020236) and NSF award (#1257101) to D.C.

Key terms

Agent A virtual agent is an autonomous character in the virtual world; a digital robot, who is not an *avatar* (see below). Rather, an agent's actions are controlled by a computer, and not by a human actor.

Avatar The character that embodies a human immersed in the virtual world; the digital persona of a human actor.

HMD Abbreviation for *Head Mounted Display*. A helmet containing the video screen on which an iVR participant views the virtual world.

iVR Abbreviation for *Immersive Virtual Reality*. The kind of virtual reality system in which percepts in the visual modality (and sometimes other sensory modalities as well) are entirely determined by the virtual environment; participants have no access to the real (visual) world, and are therefore *immersed* in the virtual world.

Presence A participant's subjective sense of immersion in the virtual world.

Uncanny Valley A region of the continuum between artificial-looking and real-looking stimuli. People's level of comfort interacting with robots (physical or virtual) generally increases as the robots' appearance becomes more realistic; an exception to this trend, however, is that people often feel uncomfortable with robots or other devices that look about 90% (but not entirely) lifelike. These devices are said to fall into the *uncanny valley.*

References

Asimov, I. (1951). *Foundation*. New York: Doubleday.

Babel, M. E. *Phonetic and social selectivity in speech accommodation*. PhD dissertation, University of California, Berkeley, 2009.

Baum, L. F. (1958). *The wizard of oz.* New York: Scholastic.

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, *13*, 103-124.

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

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

Brugman, H., & Russel, A. (2004). Annotating multimedia/multi-modal resources with ELAN. In *Proceedings of LREC 2004*, Fourth International Conference on Language Resources and Evaluation.

Chang, C.-Y., Lange, B., Zhang, M., Koenig, S., Requejo, P., Somboon, N., Sawchuk, A. A., & Rizzo, A. A. (2012). Towards pervasive physical rehabilitation using Microsoft Kinect. In *6th International Conference on Pervasive Computing Technologies for Healthcare*, 159-162. IEEE.

Chu, M., & Hagoort, P. (2014). Synchronization of speech and Gesture: Evidence for interaction in action. *Journal of Experimental Psychology: General*, *143*, 1726.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

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

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

Fox, J., Arena, D., & Bailenson, J. N. (2009). Virtual reality: A survival guide for the social scientist. *Journal of Media Psychology*, *21*, 95-113.

Fox, J., Bailenson, J. N., & Tricase, L. (2013). The embodiment of sexualized virtual selves: The Proteus effect and experiences of self-objectification via avatars. *Computers in Human Behavior*, *29*, 930-938.

Friedman, D., Steed, A., & Slater, M. (2007). Spatial social behavior in second life. In C. Pelachaud, J. C. Martin, E. Andre, G. Chollet, K. Karpouzis, & D. Pelé (Eds.), *Proceedings of*

Intelligent Virtual Agents: 7th International Working Conference (pp. 252-263). Berlin: Springer.

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

Gijssels, T., Staum Casasanto, L., Jasmin, K., Hagoort, P., & Casasanto, D. (2016). Speech accommodation without priming: The case of pitch. *Discourse Processes, 53*, 233-251.

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context and consequences. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of Accommodation* (pp. 1-68). Cambridge & Paris: Cambridge University Press & Editions de la Maison des Sciences de l'Homme.

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.

Groom, V., Bailenson, J. N., & Nass C. 2009. The influence of racial embodiment on racial bias in immersive virtual environments. *Social Influence*, *4*, 231-248.

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

Haviland, J. B. (1993). Anchoring, iconicity, and orientation in Guugu Yimithirr pointing gestures. *Journal of Linguistic Anthropology*, *3*,3-45.

Heeter, C. (1992). Being there: The subjective experience of presence. *Presence: Teleoperators and Virtual Environments, 1*, 262-271.

Heyselaar, E., Hagoort, P., & Segaert, K. (2015). In dialogue with an avatar, language behavior is identical to dialogue with a human partner. *Behavior Research Methods*, 1-15.

Jaeger, T. F., & Snider, N. (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 Conference* (pp. 1061-1066). Austin, TX: Cognitive Science Society.

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

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.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science, 11*, 32-38.

Levelt, W. J. M., Richardson, G., & La Heij, W. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language*, *2*, 133–164.

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences*, *8*, 108-114.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

Meadows, M. S. (2007). *I, avatar: The culture and consequences of having a second life*. New Riders.

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

Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., Gaggioli, A., Botella, C., & Alcañiz, M. (2007). Affective interactions using virtual reality: The link between presence and emotions. *CyberPsychology & Behavior*, *10*, 45-56.

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.

Yee, N., & Bailenson, J. (2007). The Proteus effect: The effect of transformed self-representation on behavior. *Human Communication Research*, *33*, 271-290.

Van Berkum, J. J. A., De Goede, D., Van Alphen, P. M., Mulder, E. R., & Kerstholt, J. H. (2013). How robust is the language architecture? The case of mood. *Frontiers in Psychology*, *4*, 505.

Verne, J. (1870). *20,000 leagues under the sea*. Translated by Anthony Bonner, 1962. New York: Bantam.

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

Further reading and resources

Loomis, J. M., Blascovich, J. J., & Beall, A. C. (1999). Immersive virtual environment technology as a basic research tool in psychology. *Behavior Research Methods, Instruments, & Computers, 31*, 557-564.

McCall, C., & Blascovich, J. (2009). How, when, and why to use digital experimental virtual environments to study social behavior. *Social and Personality Psychology Compass, 3*, 744-758.

Tarr, M. J., Warren, W. H. (2002). Virtual reality in behavioral neuroscience and beyond. *Nature Neuroscience, 5*, 1089-1092.