This is the accepted version of the manuscript in:
***Research Methods in Applied Linguistics*.**
DOI: https://doi.org/10.1016/j.rmal.2022.100029
URL: https://www.sciencedirect.com/science/article/pii/S277276612200026X

DISCOURSE COMPLETION TASKS MEET VIRTUAL REALITY: A PILOT STUDY ON VIRTUAL REALITY AS AN ELICITATION INSTRUMENT

Mason A. Wirtz[1]
[1] *University of Salzburg*

Corresponding Author
Mason A. Wirtz
mason.wirtz@stud.sbg.ac.at
University of Salzburg
Department of German Language and Literatures
Erzabt-Klotz-Straße 1
5020 Salzburg

ABSTRACT

In the interdisciplinary field of applied linguistics, research on the acquisition of sociolinguistic competence in the second language is growing quickly, yet comparatively little effort has been dedicated to developing innovative data collection instruments that combat long-known methodological shortcomings. This pilot study, based on data from native German speakers from Austria, evaluates the use of virtual reality (VR) as a means to (a) compensate for weaknesses of the sociolinguistic/semi-directed interview in sociolinguistic research and (b) elicit mechanisms of linguistic convergence towards a respective interlocutor's language variety. Qualitative content analysis showed high rates of immersiveness in interaction with the virtual interlocutors. The production data showed clear between-participant tendencies to converge to the respective virtual interlocutor's Austrian dialect or standard German variety. These findings are discussed against the backdrop of L2 sociolinguistic acquisition research and highlight the necessity for methodological advances in data collection practices.

Keywords: virtual reality; linguistic accommodation; sociolinguistic interview; second language research methodology

INTRODUCTION

Sociolinguistic competence involves "the acquisition and fine-tuning of the variation patterns which are used by all speakers of their language or languages" and "the ability to style-shift and use registers appropriately" (Regan, 2010, p. 22). While many resources have been harnessed to explore how second language (L2) learners acquire sociolinguistic competence, comparatively little effort has been dedicated to developing innovative instruments for collecting data on L2 sociolinguistic competence. The overwhelming majority of research projects investigating L2 sociolinguistic acquisition have made use of Labovian (e.g., Labov, 1972a) sociolinguistic/semi-directed interviews. These sociolinguistic interviews have been employed to elicit numerous L2 structures/phenomena, which have then been used to operationally define sociolinguistic competence (in German, for example, the variable use of standard German vs. Swiss Alemannic dialect [Ender, 2019]). While the sociolinguistic interview represents a reasonable starting point for projects examining L2 sociolinguistic competence, it has also seen much criticism. Among such criticisms, sociolinguistic interviews have been charged with the inability to control for power and interactional asymmetries between dyads and to provide the same degree of formality across interview dyads/groups (see e.g., Schilling-Estes, 2008). There thus exist continued calls in L2 sociolinguistic acquisition research to "expand methodologically beyond what currently exists" (Geeslin, 2010, p. 514).

In response to such calls, the present pilot study aims to introduce and evaluate virtual reality (VR) as a new elicitation instrument to collect data on L2 sociolinguistic competence. The current study follows one of Ender's (2019) operationalizations of sociolinguistic competence, namely subjects' linguistic convergence towards the language variety of the respective standard German-speaking or dialect-speaking interlocutor. To collect data on this phenomenon, I draw on the methodological innovation and unique potential of a VR-based elicitation instrument. VR as an elicitation instrument presents itself as a viable alternative to previous elicitation methods (such as semi-directed/sociolinguistic interviews) used in L2 sociolinguistic acquisition research and is advantageous in several regards: A VR design allows researchers to maintain experimental control over (a) stimulus choice and design, (b) power and interactional asymmetries between dyads and (c) the degree of formality. Moreover, VR eliminates the possibility of unintentional interviewer accommodation to the interviewee—a substantial drawback in sociolinguistic interviews, particularly with L2 interviewees—and reduces the artificial spatial divide between participant and stimulus classic to traditional experimental procedures. Whereas the aforementioned advantages can be ensured by more experimental designs, such designs typically come at the cost of ecological validity. VR introduces the unique potential to reconcile experimental control and ecological validity (Peeters, 2019) by subjecting participants to a highly immersive virtual setting. Given VR's high degree of immersiveness and experimental flexibility, it is possible to compensate for many known weaknesses of previous methods (e.g., the sociolinguistic interview) and collect realistic speech data in an experimentally controlled yet immersive setting that does justice to the "multimodal richness and dynamics of everyday communication" and interaction (Peeters, 2019, p. 899).

The article begins with a methodological discussion on the sociolinguistic interview and several of its shortcomings. Following, the *virtual reality discourse completion task* is introduced, and it is elucidated how VR can compensate for known weaknesses in the (Labovian) sociolinguistic/semi-directed interview method. Using data from nine native German speakers from Austria, the present study sets out to evaluate virtual reality's unique methodological potential for the fields of variationist SLA and sociolinguistics, focusing specifically on the VR-based instrument's immersiveness and ability to elicit mechanisms of linguistic convergence. The article concludes with a renewed call for VR and other

methodological innovations that can help researchers compensate for the various weaknesses of our research instruments.

BACKGROUND

The Sociolinguistic Interview

The Labovian sociolinguistic interview is a method developed for Labov's (1966) study of New York City's Lower East Side and remains ideologically central to the field of sociolinguistics. The term "interview" is a misnomer, however, in that the goal is less to interview subjects but rather to have an "informal chat" to elicit conversational speech (Milroy & Gordon, 2003; Tagliamonte, 2006). The traditional sociolinguistic interview comprises several subsections designed to elicit speech from an individual and is structured to move temporally across various contextual language styles such as casual, careful, reading, word list and minimal pair (Labov, 1972a). This methodological approach was conceptualized to (a) generate insights into speakers' patterns of stylistic variation and subsequently (b) identify each speaker's most casual speech style, i.e., the vernacular. The rationale for the focus on the vernacular is outlined in Labov's Vernacular Principle, which asserts that "the style which is most regular in its structure and in its relation to the evolution of language is the vernacular, in which the minimum attention is paid to speech" (Labov, 1972b, p. 112). As such, a main goal of conducting sociolinguistic interviews is to minimize speakers' attention to their speech and thereby mitigate the effects of the well-known Observer's Paradox. As Labov (1972b, p. 113) explains, "To obtain the data most important for linguistic theory, we have to observe how people speak when they are not being observed."

The sociolinguistic interview and its adaptations have not gone without criticism. For example, Wolfson (1976) argued that the sociolinguistic interview situation does not provide a *more* natural situation, but rather a *less* natural one, given that interviewees typically expect face-to-face interviews to be formal in nature and follow a predetermined set of questions. It has also been argued that, even if the interviewer relinquishes control to the interviewee by allowing the interviewee to take charge of the conversation, there remain inherent power asymmetries, such that the interviewer typically "holds the more powerful conversational role" (Schilling-Estes, 2008, p. 974). Milroy and Gordon (2003, p. 62) argue that turn-taking rights "are not equally distributed as they are in conversational interaction between peers," because the interviewer controls the discourse of the interview regarding topics and the form of questions. A further major critique is the presupposition that "with minimal intervention of the researcher the interview could become as similar as possible to a 'natural conversation'," leading to the interview being understood "more as a context to be erased than as an interactional event" (de Fina & Perrino, 2011, p. 4). By focusing only on the speech of the interviewee, especially by analyzing it independently of the interviewer/interlocutor, researchers neglect the emergent qualities of the speech shaped by both interactants. Viewed against the backdrop of Accommodation Theory (Giles et al., 1991), for example, we can see that attempting to erase the interview context as an interactional event and strictly focus on the interviewee's speech poses particular problems, such that interviewers may adjust and shift their accent and other prosodic, discursive, and syntactic features as a function of the interviewee's speech, identity, and role in the conversation.

Given such shortcomings in sociolinguistic interviews, a goal of sociolinguistic research has been to provide (partial) solutions to these criticisms by modifying the dynamics of the interview situation, e.g., by (a) migrating away from the one-to-one format, (b) placing pairs of speakers in a room and allowing for free conversation, (c) providing more interviewers/interviewees, or (d) conducting interviews in friendship pairs (Milroy & Gordon, 2003). Yet, these adaptions carry with them many of the same challenges previously

mentioned. Even with the numerous variations of the sociolinguistic/semi-directed interview, it is difficult to control for, inter alia, the level of formality, the social parameters and power distributions among the interviewer(s) and interviewee(s) (Milroy & Gordon, 2003), the Observer's Paradox, and the level of interviewer accommodation to the interviewee across interviews.

The Sociolinguistic Interview in L2 Sociolinguistic Acquisition Research

Since the late twentieth century, variationist sociolinguistics and second language acquisition (SLA) research have joined forces to investigate the extent to which L2 learners acquire sociolinguistic variation commonly found in native speaker speech. In L2 sociolinguistic acquisition research, (adaptations of) the sociolinguistic interview have been a predominant form of data collection and are typically featured as a form of oral production task, the goal being to provide a communicative setting to elicit informal speech from L2 learners. However, the prevalent sociolinguistic and semi-directed interview methods used in L2 variation acquisition research are likewise affected by the methodological shortcomings addressed in the previous section. It is moreover plausible to argue that several of the methodological shortcomings are exacerbated in interviews with L2 learners, particularly with respect to accommodation effects. For example, self-reports in Drljača Margić (2017) indicated that native English speakers were highly inclined to accommodate to the speech of non-native English speakers. Results in Ender and Kaiser (2009) similarly indicated that native speakers in Bavarian-speaking Austria typically use standard German rather than the local dialect when speaking with L2 German learners. Vergeiner (2020) likewise underscores this finding in his analysis of L1 speakers addressing L2 speakers, noting that native speakers drew predominately on standard German speech and tended to avoid dialectal utterances. Given that accommodation effects have been posited to be both socially motivated to enhance communicative success (Giles et al., 1991) and an involuntary/unconscious phenomenon (Pickering & Garrod, 2004), it is highly likely that interviewers accommodate to the speech of L2 interviewees with respect to the language variety (e.g., standard versus nonstandard) and/or other phonetic and prosodic features, though likely to different extents across interviewees as a function of, for instance, proficiency.

Ender (2019) explored accommodation by L2 learners to the standard German and Swiss dialect variety of the respective two native speaker interviewers and found that select L2 learners did systematically switch varieties to match the native speaker interlocutor. Dewaele (2004a) similarly found convergence effects in his sample, in that L2 learners used informal variants in intended informal interactions to maximally converge to the native speakers' variation patterns, though these convergence effects were limited to a single variant (*ne* omission) and did not hold across further variants. A point to ponder is whether and to what extent unintentional accommodation by the native speaker interviewers to the L2 learners (e.g., in terms of accent and other prosodic and discursive features), as well as possible power asymmetries and the learners' awareness of being observed, may affect the degree of accommodation by the L2 interviewees.

In light of the aforementioned methodological and practical shortcomings of sociolinguistic/semi-directed interviews, there is a need for innovative designs that are less susceptible to insurmountable power asymmetries and the Observer's Paradox, that can collect data across myriad contexts and can prevent accommodation by the interviewer to the interviewee. Designs of this nature are, however, scarce. As such, methodological approaches aiming to measure L2 sociolinguistic competence must innovatively tackle the challenge of developing more fine-tuned and reproducible elicitation instruments which measure the construct interpretations proposed for sociolinguistic competence in the relevant research design. Virtual reality methodologies present a viable solution to many of these problems, in

that they (a) afford high experimental control without sacrificing ecological validity, (b) allow for reproducible science across participants and labs, and (c) can be designed to elicit any number of variable structures and phenomena that can be used to operationalize sociolinguistic competence. One such phenomenon which can be elicited with VR and used to operationalize sociolinguistic competence is *linguistic convergence*, the rationale for which will be discussed in the following section.

Behavior of Interest: Linguistic Convergence

No single task or measurement can capture sociolinguistic competence in its entirety. However, tasks extracting learners' ability to (a) (consciously or otherwise) observe the social dimensions of interaction and (b) act on these observations by producing intra-speaker, sociolinguistically motivated variation might better capture the multidimensionality of sociolinguistic competence. For this reason, a novel virtual reality instrument aiming to elicit mechanisms of linguistic convergence was designed, as linguistic convergence presents a meaningful way to investigate the interplay of productive and interpretive abilities. Regan (2010, p. 24) underscores that L2 speakers "want to accommodate to their interlocutors, to fit in with the new group/community" and that "[s]ociolinguistic competence is particularly salient in terms of "fitting in." It thus stands to reason that linguistic convergence is a robust measure of sociolinguistic competence. Before introducing the novel virtual reality instrument developed to elicit linguistic convergence, I outline the advantages of using linguistic convergence as an operational measure of sociolinguistic competence in more detail.

On the whole, speakers align their linguistic behavior to that of their interlocutor. For example, the Bavarian-Austrian variation landscape (the focus in this study; for a brief overview, see Ender and Kaiser [2009]) is oftentimes conceptualized along the lines of a linguistic continuum, meaning speakers command two "end poles," namely a local dialect and (Austrian) standard German, as well as a plethora of intermediate forms between the poles. Against this backdrop, linguistic convergence can be understood as, e.g., reducing the use of dialect in favor of standard German when interacting with a standard German speaker, and vice versa. Yet, while convergence is a commonly cited phenomenon when modelling patterns of variation, the motivation behind it remains a matter of debate. This question of *why* has been addressed repeatedly and under numerous theoretical paradigms, e.g., "Interactive Alignment" (Pickering & Garrod, 2004) and "Communication Accommodation Theory" (Giles et al., 1991), among others. Taken together, these approaches highlight linguistic convergence as a notion of cooperativity, in which speakers—not necessarily consciously—reduce salient linguistic differences in terms of, e.g., accent or dialect to sound closer to that of the interlocutor. Among many, e.g., Coupland (1985) posits that the level of convergence in natural conversational interactions is mediated by speakers' interactional goals, i.e., to make an interlocutor (intuitively) think, do, or feel something. Conversely, mechanistic theories of dialogue such as *interactive alignment* (Pickering & Garrod, 2004) question whether immediate social motivations mediate adaptive behavior in communication, positing rather that speakers automatically align their use of language to that of the interlocutor. There is a body of growing evidence which suggests that the interplay of these theories might be a better representation of convergence/accommodative behavior, in that overlearned social motivations and automatic alignment processes may mediate and/or motivate accommodation in the long run (e.g., Casasanto et al., 2010).

I argue that linguistic convergence meets the multidimensional measurement demands of sociolinguistic competence, in that it requires speakers to (a) observe and replicate "the details of the variability present in the native speaker system" (Regan, 1996, p. 178) to subsequently (b) make variable use of sociolinguistic variables (Dewaele, 2002), which can

all be used to measure speakers' ability to (c) "style-shift consistently and appropriately" (Dewaele, 2004a, p. 432). Thus, linguistic convergence capitalizes on not only speakers' productive abilities, but their interpretive ones as well. Rather than assessing sociolinguistic competence on the basis of frequency-based measurements of singled-out sociolinguistic variants, linguistic convergence offers a broader, more global measure, in that its measurement can include a plethora of sociolinguistic features (e.g., phonological, morpho-syntactic, lexical). On the whole, measuring linguistic convergence can be argued as advantageous when compared to measuring singled-out variants, because: (a) it is oftentimes time consuming to collect such extensive free speech data that a wide variety of sociolinguistic variables can be individually analyzed; (b) some sociolinguistic variables might constitute only a small portion of a speaker's repertoire; and (c) the variable use of one sociolinguistic variable does not necessarily ensure the variable use of another (Dewaele, 2004a, 2004b), making it necessary to examine a plethora of variables for accurate measures of sociolinguistic competence. Of course, to examine linguistic convergence in this manner, the requirement is an experimental setting in which (a) the interlocutor can prevent their own linguistic repertoire from being influenced by the naïve participant's speech and, ideally, (b) the stimuli provided by the interlocutor are constant across participants, ensuring complete control over the social, power-related, and linguistic elements towards which participants converge. This is where we can place a premium on tasks exploiting the advantages of virtual reality.

Task Characteristics: The Virtual Reality Discourse Completion Task

For the investigation into speakers' inherent abilities of linguistic convergence, the aim was to conceive an environment in which interactional goals cannot be directly achieved. Methodologically speaking, the rationale behind this choice was to allow for a more nuanced understanding of speakers' overlearned convergence behavior by controlling for the confounding variables of (a) live interlocutors' own (unintentional) accommodative behavior; (b) power asymmetries between different dyads and (c) unaccountable social parameters—factors less controllable in e.g., sociolinguistic interviews and adaptations thereof. Given this objective, VR emerged as a viable option. Many projects employing VR use computer-generated virtual environments in addition to computer-partner(s) (i.e., avatars). Of course, such VR projects are not always (financially and/or technologically) feasible and there remains some skepticism towards VR in linguistics. Heyselaar et al. (2017, p. 2351) suppose that this lack of use of VR is attributed to the assumption that "humans do not interact with computers in the same way that they interact with other humans, making any behavioral measure of language interaction with a computer-partner ("avatar") ecologically equivocal." Yet, we have seen this assumption to be largely unjustified (Casasanto et al., 2010; Heyselaar et al., 2017; Nass & Moon, 2000; Peeters & Dijkstra, 2017; Stoyanchev & Stent, 2009), as speakers socially and interactionally engage with virtual/digital avatars in human-like ways.

In creating a VR elicitation instrument for collecting data on learners' ability to converge in the L2, one of my primary goals was to develop a VR-based methodology that "the average researcher" can effectively use with minimal financial burden. Thus, as opposed to developing computer-generated virtual environments and avatars, I produced VR videos with human interlocutors, which should not only reduce the time needed to create VR elicitation instruments but also allow maximal ecological validity and ensure the most realistic elicitation setting possible. Specifically, I propose combining VR with discourse completion tasks (DCTs), similar to emerging VR DCT instruments used in pragmatics research (Taguchi, 2022a, 2022b). In essence, the DCT is an instrument that can be administered orally or in written form and provides various scenarios (e.g., situations with differing degrees of social distance/dominance) designed to elicit a desired speech act. In

what follows, I will focus exclusively on a variation of the *oral dialogue construction* DCT (Bardovi-Harlig & Hartford, 1993).

In this new DCT variant, the *virtual reality oral dialogue construction discourse completion task*, participants receive social and situational information in the form of a virtual narrator as a means of explicit contextualization, after which a dialogue is initiated by the interlocutor. This fulfills a range of functions that are lost or simply not controllable in a live interview situation:

- Each participant is provided the same social-situational information, the rationale being to reduce potential power and/or social asymmetries;
- the (audio-)visual setting can actively be tailored to equate the social-situational information provided (e.g., the video could be set in a coffeehouse or an apartment);
- the well-known Observer's Paradox is reduced by the immersive environment (see Results section);
- there exists no risk of unintentional accommodation to the "interviewee" by the VR interlocutor, given that the VR interlocutors' utterances are pre-recorded;
- the researcher retains full control over the scenario with respect to pragmatic, social, contextual, and other factors;
- the design guarantees absolute repeatability of each sequence; and
- the researchers' interactional role with the participant is minimized.

While we can guarantee similar repeatability using only video or audio material, VR offers the additional "immersive" factor. For example, in Taguchi's (2022b) analysis of L1 and L2 speaker fluency in computer-based versus VR-based tasks, she found that social-interpersonal demands were higher in the VR-based tasks, such that participants were more attentive to the impact their linguistic behavior could have on the interlocutor. Taguchi (2022a) notes in a similar vein that the realistic nature of VR could have invoked stronger emotions and affective responses in participants, underscoring VR's immersiveness. As a further advantage, VR also provides a possible solution to the much thematized shortcomings of DCTs, i.e., in their inability to capture as rich or as extensive data as, e.g., roleplays or interviews (Billmyer & Varghese, 2000; Vanrell et al., 2018). While the speech data from the VR may be less "extensive" in that there are fewer hours to analyze, the data may well be of a more realistic quality and less prone to bias and distortion due to unperceivable yet very present social-situational asymmetries.

THIS STUDY

Research Design

The method in this study is twofold: In exploring the extent to which VR provides an immersive environment, qualitative data from retrospective interviews are inspected. This approach allows the investigation of participants' experience *during* the VR and should facilitate our understanding of user interaction with the VR environment. The second method in this study can be characterized as an equal-status concurrent mixed-methods design (see Johnson & Onwuegbuzie, 2004), with the basis being that of complementarity and triangulation. This mixed-method approach generates descriptive-quantitative data on participants' group-level convergence patterns and links such patterns to the qualitative, introspective data on participants' rationale behind their conscious changes in varietal behavior. Thus, this study aims not only to shed light on the between-participant convergence patterns but also to identify which (contextual) aspects of the VR environment motivated the observed trends in varietal behavior, which can provide complementary evidence from a qualitative perspective concerning how well this VR instrument elicits mechanisms of linguistic convergence.

Research Questions

The following research questions guide my study:

RQ1. To what extent does this virtual reality instrument provide an immersive, naturalistic, and communicative environment for participants?
RQ2. Does this virtual reality instrument elicit mechanisms of linguistic convergence to the respective variety of the dialect-speaking and standard German-speaking virtual interlocutor?
RQ3. To which contextual factors in the virtual reality environment and social traits of the virtual reality interlocutors do participants attribute the observed group-level patterns of varietal behavior?

RQ1 is tackled in a qualitative manner. To elicit deeper information on users' immersive experiences from an introspective angle, retrospective interviews were conducted directly after participants had completed the questionnaire to avoid "fading affect bias" (Ritchie et al., 2006).

The final two RQs deal with the instrument's effectiveness in eliciting mechanisms of linguistic convergence from a descriptive-quantitative (RQ2) and qualitative (RQ3) perspective. The goal of RQ2 is to uncover any group-level trends, whereas RQ3 deals primarily with introspective data on participants' rationale behind their respective varietal behavior in the VR. The data inspected in this RQ should act as the explanatory force behind the group-level trends pertaining to RQ2.

The overarching goal of the present contribution is to position VR as a suitable alternative for semi-directed/sociolinguistic interviews. To be considered a suitable alternative, however, this VR methodology must (a) provide an immersive, naturalistic, and communicative setting (as this is a primary goal of sociolinguistic interviews) and (b) be capable of eliciting phenomenon and variable structures that can be used to operationalize sociolinguistic competence. From complementary angles, the three RQs outlined above aim to evaluate whether this novel VR elicitation methodology fulfils the aforementioned requirements. Results pertaining to RQ1 shed light on the *immersiveness* of the virtual reality instrument. This is pertinent to the current evaluation of the VR instrument, as it offers insights as to the ecological validity of the data collected and the extent to which the VR provides a naturalistic and communicative setting. Building off this, RQ2 homes in on the VR instrument's capacity to elicit mechanisms of linguistic convergence (the chosen operational measure of sociolinguistic competence). RQ3 complements RQ2, in that it focuses on which contextual aspects of the VR environment and social traits of the virtual interlocutors motivated the observed varietal behavior; this should provide additional, complementary evidence from a qualitative perspective as to whether/how well the VR instrument elicits mechanisms of linguistic convergence.

For results on participants' user experience with the VR, see Supplementary Material.

Participants

I developed this VR elicitation instrument with L2 learners in mind, as a method to meet the multidimensional complexity of measuring sociolinguistic competence. However, L2 learners are exceedingly diverse in their L2 sociolinguistic competence. As such, they were not ideal subjects for a pilot study evaluating the effectiveness of the instrument's capacity to elicit trends of linguistic convergence. I therefore recruited nine speakers native to the Central Bavarian area in Austria (range = 24–28 years, M = 25.8 years, SD = 1.2, 7 women). Because speakers in the Bavarian-Austrian context are notorious for commanding diverse varietal

repertoires, the sample was carefully chosen. It comprised speakers with mastery of a large varietal spectrum between standard German and Austrian dialect. The rationale was to ensure the sample pool consisted of participants who unambiguously could produce both dialectal and standard-near speech. To confirm participants indeed commanded the necessary varietal repertoire, they completed a biodata questionnaire during an initial screening. The questionnaire collected information on a seven-point Likert-scale (0–6) on participants' self-reported receptive and productive proficiency in standard German (range = 5–6, $M = 5.7$, $SD = 0.43$) and dialect (range = 5.5–6, $M = 5.9$, $SD = 0.2$).

Tasks and Procedure

Virtual reality discourse completion task. The data on linguistic convergence were collected by means of a VR headset. This configuration was based on the VR Shinecon (model: FIYAPOO) headset for smartphones. The smartphone was an iPhone 11 with a 6.06" full HD screen, 4 GB RAM, and a gyroscope sensor. The iPhone ran the software version 14.4.2 and used the pro-version VRPlayer app configured with a special profile created for a headset-smartphone combination. The field of view was ~180° with a refresh rate of 60Hz. While previous virtual reality studies have made use of other, more advanced hardware configurations, research has shown that these more advanced/expensive hardware configurations do not produce significantly different results than low-cost VR headsets for smartphones (Amin et al., 2016; Papachristos et al., 2017).

Participants were subjected to two sets of VR discourse completion tasks. In each task, the subjects interacted with two different interlocutors speaking respectively either standard German or dialect. The standard German-speaking interlocutor was 24 at the time of filming, and the dialect-speaking interlocutor was 27, in line with the mean age of the subject pool. It was hoped that this would encourage participants to consider the interlocutors as peers.

Both VR sets were informal in nature (VR set 1: running into a friend in the hall; VR set 2: drinking coffee at an outside café). To ensure realistic utterances by the VR interlocutors, the VR interlocutors were given a list of topics (e.g., when should I use dialect versus standard German) in a preparation meeting and asked which questions they could imagine themselves asking in everyday situations. The VR interlocutors were then asked in what type of situation, using what phrasing and in which variety they could imagine themselves asking questions about the respective topics. For example, one VR interlocutor responded that she could imagine asking questions in dialect about when to use dialect versus standard German when running into a friend in a hallway or similar setting, given that she grew up in a dialect-speaking environment in another province in Austria. During the recordings, the VR interlocutors were given their specified topics and list of questions they could ask. The "director" of the VR videos was a close acquaintance of both VR interlocutors, and the VR interlocutors were instructed to ask the director questions on the topics as if talking to him in everyday life to best ensure informal utterances.

The experimental procedure was as follows:

- The participants were first shown an introductory VR photo (59 seconds) which introduced them to the experiment and familiarized them with the headset.
- Directly following, participants were subject to a second introductory VR environment (1 minute, 31 seconds) which they could freely explore. The rationale was to provide participants with enough time to fully accustom themselves to the VR headset before beginning the official experiment.
- Afterwards, participants were immersed in the VR oral dialogue construction discourse completion tasks.

Each VR set began with explicit contextual information read in English to (a) preclude chances of accidental accommodation to instructions provided in (standard) German; (b) ensure participant understanding of the social and situational context notes and (c) reduce risks of power asymmetries. During the tasks, a standard German-speaking and dialect-speaking interlocutor asked the participants a series of conversational questions. It must, however, be borne in mind that the VR sets were interactionally a little different. In the first VR set, participants spoke sequentially with the interlocutors—first with the dialect-speaking interlocutor, then with the standard German-speaking interlocutor. In the second VR set, participants first interacted with the standard German interlocutor, after which the dialect speaker joined the standard German-speaking interlocutor at a table. While participants were sensible to this social change, the data revealed similar convergence trends across both VR sets (see online supplementary material, https://osf.io/ebcdf/).

On a final note, it is important to note that convergence is realized in interaction and negotiation with all present interlocutors. In this given case, however, the VR interlocutors cannot engage in any social negotiation of varietal choice, and subjects similarly cannot assess the degree to which varietal adjustments facilitate communication. As such, the combination of VR with DCTs is strictly a means to evaluate subjects' (theoretical) capacity to engage in linguistic convergence in live interactions based on their degree of convergence when constructing dialogue with the VR interlocutors in a controlled experimental setting. While such a design does not allow for live interaction, it does facilitate experimental control of all stimuli, allowing for more consistent comparisons across both tasks and subjects.

Retrospective interview. After the virtual reality tasks and questionnaire, information was gathered about the following topics via retrospective interviews: (a) how participants felt they handled the virtual reality situations, (b) their reflections on conversing with the virtual interlocutors, (c) possible changes to the VR environment, and (d) notes on participants' overall satisfaction with the VR experience.

Data Analysis

Qualitative analysis. For the qualitative analysis of the retrospective interviews (RQ2), I pursued an inductive, bottom-up approach and performed a content analysis in MAXQDA (http://www.maxqda.com/) using the methods outlined in Braun and Clarke (2006).

Coding the speech data. In preparing the speech data collected in the VR discourse completion tasks for the quantitative analysis (RQ3), I first segmented all utterances into TCUs ("turn-constructional units" [Schegloff , 1996]) using ELAN (ELAN, 2021). A TCU is a "potentially complete turn" (Schegloff, 1996, p. 55; Selting, 2000, p. 480). As such, it is considered to be an "interactionally relevant unit" (Kaiser, 2021, p. 9). The advantages in working with TCUs in this context are outlined extensively in Kaiser (2021) and Kaiser and Ender (2021). Long turns with complex syntactic constructions were segmented into multiple TCUs when syntactic and prosodic (e.g., pauses, rephrases) features unambiguously indicated that additional psycholinguistic planning processes took place.

After the utterances had been segmented, the TCUs were transcribed orthographically, such that a distinction between dialectal and standard-like features was apparent. The transcribed data were then imported into MAXQDA. Each word was coded as (a) "standard language" (features consistent with standard language use, i.e., *Gebrauchsstandard* [Kleiner & Knöbl, 2015]); (b) "dialect" (features consistent with Central Bavarian-Austrian dialect; Ender & Kaiser, 2009) or (c) "mixture" (with words that [a] were not unambiguously standard German or dialect or [b] consisted of both dialect and standard German features). On the basis of the word-level coding, the TCUs were coded as (a) "standard language" (standard and mixture words), (b) "dialect" (dialect and mixture words), or (c) "mixture"

(standard, dialect, and mixture words; a mixture of standard and dialect words or simply mixture words).

RESULTS

RQ1: On the Immersiveness of the Virtual Reality Instrument

To begin, the qualitative data were analyzed in order to outline participants' experiences from an introspective angle. Participants referenced overall the following four phenomena:

- A comparatively high sense of spatial detachment, such as suppressing the setting within the objective reality;
- The realistic involvement in the VR conversations with the interlocutors;
- A lack of total spatial dissociation, such as retaining an underlying knowledge of the objective reality;
- Emotional involvement, such as a sense of initial nervousness, quickly replaced by sensations of relaxation, humor, comfort, etc.

The participants collectively underscored their sense of spatial detachment, in that they were able to suppress the external experimental setting: "It was pretty easy to suppress the fact that you [the researcher] were sitting there" (Lusia, 01:05). Participants also remarked about their absent sense of "being in an experiment": "I didn't even think that this is actually an experiment" (Lili, 01:35). Their sense of spatial immersion in the VR environment also appeared to be high: "I had the feeling that [NAME, the virtual interlocutor] was standing next to me" (Leonie, 01:24). As for optic factors, the "visual transportation" to the VR environment was repeatedly mentioned: "[F]rom the looks of it, it really looks like when you are sitting at the [PLACE]" (Luisa, 01:56).

Participants did not appear to have difficulty conversing with the virtual interlocutors. In fact, an overall sense of immediate, realistic involvement in the VR conversations seemed to be the case: "I was very positively surprised by how much you get 'pulled into the train' [German idiom, meaning 'sucked into the situation'] and how much you get into this situation" (Debora, 00:12). Moreover, it was recorded on multiple occasions that participants felt as if they were having an actual conversation: "I felt as if I was in an actual conversation and I didn't just say sentence after sentence. It was a natural conversation" (Lars, 00:40). Participants also commented on the lack of "staging" or artificiality: "It [the conversation] happened very naturally in the environment, as if you weren't somehow staged, but it was rather a real conversation in this situation" (Lars, 01:05).

As natural as the experience was for these participants, it should not be construed as total spatial dissociation; participants retained an underlying knowledge of the objective reality, though to a rather subdued extent: "It [the VR] was good enough and convincing enough that I didn't […] consciously […] think it wasn't real. […] [I]t wasn't as if now I thought I happened to be somewhere else—more of an in-between" (Leo, 00:20). Even in retaining knowledge of the objective reality, VR still goes a long way in providing an immersive elicitation environment: "Sure, you know that it's not real, but it's [VR] a big help—it takes away your inhibitions" (Lili, 00:52).

Finally, while participants reported initial feelings of nervousness, these quickly faded and were replaced with sensations of relaxation, humor, and even pleasure: "It was quite exciting, even a bit fun" (Lusia, 00:41). Other participants reported no signs of nervousness: "I felt good, I wasn't nervous at all" (Freia, 00:32). Moreover, participants mentioned the overall pleasant and informal atmosphere: "I actually really enjoyed it—it's something

friendly, something light" (Leonie, 01:24) and "[t]he situation was just so friendly" (Lars, 02:02).

In sum, all study participants appear to have experienced an immersive, informal, and friendly environment that reflects/simulates a natural conversation. This speaks to (a) the VR instrument's effectiveness in eliciting realistic (informal) linguistic behaviors and thus (b) the validity of the elicited language data to be analyzed in the following section. This is a feat that the Labovian sociolinguistic interview, so it is criticized, has not yet been able to completely achieve, as it is often regarded as "*less* rather than *more* natural than more typical types of interviews, since people expect interviews to be relatively formal and to follow a set questionnaire" (Schilling-Estes, 2008, p. 974). In creating informal, immersive, and "friendly" sets of VR situations, natural speech can more easily be collected while reducing the threat of the Observer's Paradox. Even though participants did not entirely forget that they were partaking in an experiment, the sensation of "being interviewed" was reduced to such an extent that participants mention not consciously considering it. This ensures that, to the largest extent of my instrument's capacity, the confounding effects of the Observer's Paradox and the researcher's interactional role are minimized. On a last note, the fact that participants reported feeling a sense of "fun" and "tried something new" during the experiment might inspire other participants with more diverse language-biographical backgrounds to participate, perhaps even from populations who traditionally participate less in (sociolinguistic) research.

RQ2: Group-Level Trends of Linguistic Convergence in the Virtual Reality Instrument

As a whole, the speech data suggest a between-subject trend to converge to the variety of the interlocutor. This is immediately evident in Excerpt 1 when Leo produced a range of typical Bavarian-Austrian dialect utterances (e.g., *a*-darkening: *jo* rather than standard German *ja*; l-vocalization in dialect *oiso* instead of standard German *also*; the Bavarian negation particle *ned* compared to standard German *nicht*). This TCU was classified as dialect since no variant for which both a standard German and dialect realization was possible was produced in standard German.

Excerpt 1

VR Interlocutor (dialect)	Hey, guad dasst do bist! Um, du i hob a Frog an di (..). Also i bin jo vom Lond, und wir redn eigentlich NUR Dialekt miteinanda. (..) Und, in Soizburg (…) is des gonz onders. Und i wa:s einfach ned (…) Wonn soi i Hochdeitsch redn, und wonn soi i Dialekt redn? Waßt du des?
Engl.	*[Hey, good that you're here! Um, I have a question for you (..). Well, I'm from the countryside and we really speak ONLY dialect with each other. (...) And, in Salzburg (..), it is really different. And I just don't know (...) when should I speak High German, and when should I speak dialect. Do you know?]*
Leo (04:09)	Jo oiso des is eigentlich goa ned so schwa (.) weil so oft braucht mo des hochdeitsch (.) bei uns donn a wieder ned
Engl.	*[Yes, well, this isn't all that hard, because you really don't need high German all that often around here]*

With the standard German interlocutor, Leo adjusted his varietal behavior. This is reflected in Except 2, which shows that Leo did not produce several of the dialect variants seen in Excerpt 1: Leo abstains from the dialect *a*-darkening (standard German *ja* and *da* compared to dialect *jo* and *do*), *-ch* reduction (standard German *auch* instead of dialect *a*) and

the production of the dialect definite article *des*, producing rather the standard German *das*, to name a few.

Excerpt 2

VR Interlocutor (standard)	Gibt es eigentlich Orte die ich vermeiden sollte (.) weil die Leute dort einen so starken Dialekt sprechen?
Engl.	*[Are there actually places that I should avoid (.) because the people there speak such a strong dialect?]*
Leo (06:44)	Ähm (.) ja (..) wahrscheinlich (…) sicherlich (.) je weiter (..) ähm ins gebirge und je weiter westlich man kommt (.) das heißt (.) äh in pinzgau ähm und da ganz im westen und vor allem in lunggau ist es auch (.) schwer zu verstehen
Engl.	*[Um (.) yes (..) probably (…) certainty (.) the further (..) um into the mountains and the further west you get (.) that means (.) um in pinzgau um and far in the west and above all in lunggau it es also (.) difficult to understand]*

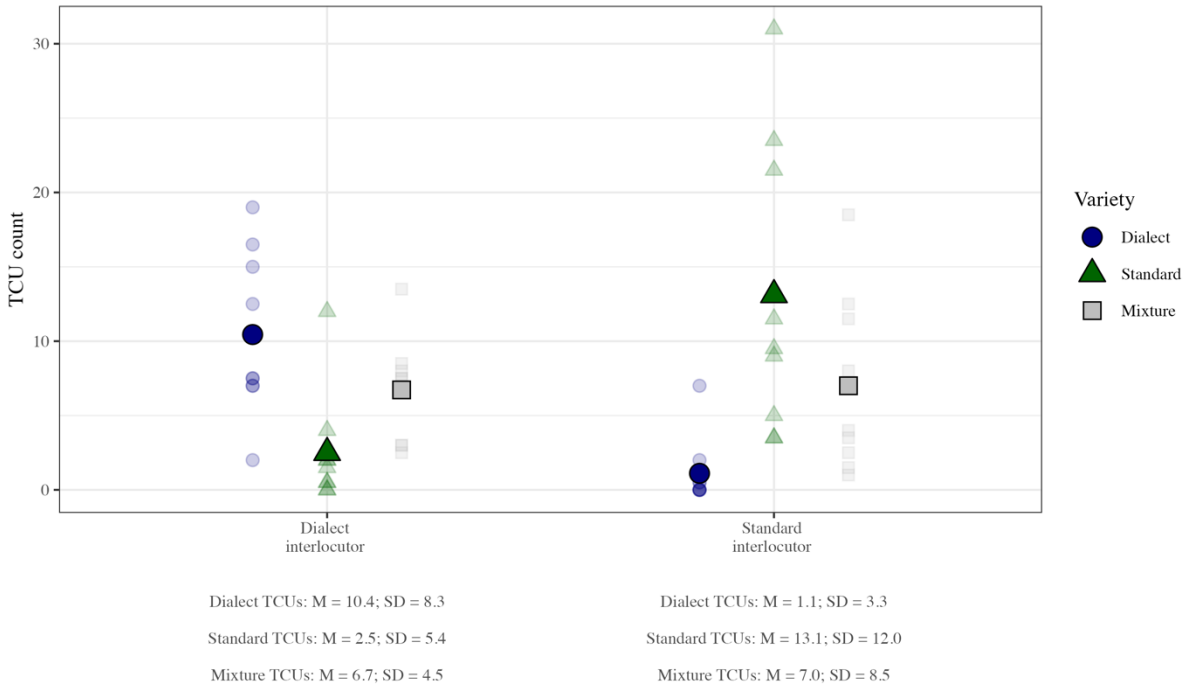
Other subjects showed more variability in their convergence tendencies, e.g., Luisa in her response to the same question the standard German VR interlocutor asked in Excerpt 2. Excerpt 3 shows that, while she did use standard German variants (e.g., standard German *sich* instead of dialect *si*; standard *Leute* as opposed to dialect *Leit*; standard *keinen* compared to dialect *koa(n)*), select dialect variants were also present in her speech with the standard German interlocutor, particularly fricative reduction (e.g., dialect *i* and *a* as opposed to standard German *ich* and *auch*).

Excerpt 3

Luisa (04:24)	Vermeiden (.) würd i (..) keinen ort (.) nur weil jemand dort dialekt spricht (…) grundsätzlich kinnan sich glaub i (..) leute a an's gegenüber anpassen
Engl.	*[I wouldn't avoid any place (.) just because someone speaks dialect there (…) in general I think the people can adapt to their counterpart]*

Figure 1

Descriptive data on the overall TCU count across varieties and interlocutors



Note. Descriptive data on the overall TCU count average in the three varieties (standard German, dialect, mixture) with the dialect and standard German interlocutor respectively (solid shapes) and each participant's TCU count in the three varieties during their virtual reality interactions with the two interlocutors (semi-transparent shapes). For an overview of the relative amount of TCUs produced with each interlocutor and participants' individual varietal behavior in the VR sets, see the supplementary figures 3 and 4 respectively in the online supplementary material.

Figure 1 displays each participant's mean TCU count in the three varieties with the dialect-speaking and standard German-speaking interlocutor respectively (semi-transparent shapes), while the solid shapes indicate the respective group means. As for the mixture variety, there does not appear to be any underlying pattern(s) to which the participants collectively adhere, given that the average mixture TCU count was similar between interlocutors. Visual inspection does, however, suggest a between-participant trend to converge towards the respective interlocutor's language variety. Yet, there is extensive variability in the data, particularly with respect to the participants' inter-individual use of the variety spoken by the respective interlocutor.

As a whole, the results illustrate that dialect TCUs were produced more in interaction with the dialect interlocutor (M = 10.4 TCUs, SD = 8.3) than with the standard German speaker (M = 1.1 TCUs; SD = 3.3). We find similar patterns when addressing subjects' variable use of standard German. In interaction with the standard German-speaking interlocutor, participants produced on average 13 TCUs (SD = 12) and reduced their standard German speech in interaction with the dialect-speaking interlocutor (M = 2.5 TCUs; SD = 5.4). Despite clear trends this descriptive-quantitative analysis provides, the standard deviations cover a large range of values, indicating a healthy amount of individual variability—a larger sample pool and/or more data from the speakers would be necessary to constrain our uncertainty of the respective differences.

For a Bayesian mixed-effects analysis of the between-subject trends, see Supplementary Material.

RQ3: Explanatory Force Behind Group-Level Trends of Convergence in the Virtual Reality Instrument

Even given the clear between-participant trends of convergence illustrated in Figure 1, the high variability leaves much room for uncertainty. For this reason, and in the spirit of complementarity, the qualitative data were inspected to find out retrospectively what participants thought while in interaction with the different interlocutors and so to ascertain whether the between-subject trends were, in fact, attributable to mechanisms of linguistic convergence. Such an integrative approach allows an elaboration of the descriptive-quantitative results from the qualitative dimension, which may help account for and/or combat the high amount of variability in the data.

The qualitative, retrospective data reveal that the respondents often attributed the changes in their varietal behavior directly to processes of convergence, particularly in converging towards the standard German variety. It is no secret that many Austrian speakers prefer their native dialect (Ender & Kaiser, 2009). Thus, the participants were not necessarily required to "change" their everyday language in interacting with the dialect-speaking interlocutor: "[W]hen I spoke dialect, it just came so naturally [in the VR]" (Lars, 01:37) and "[i]f the other person speaks a similar or a strong dialect, then you reflect that, because that's the most convenient" (Leo, 01:15). Rather, participants mentioned processes of convergence relative to the standard German interlocutor. In fact, it was noted that such tendencies of convergence towards the standard German variety were, in part, socially motivated, particularly to ensure understanding: "Since [NAME] spoke standard German to me, I spoke back in standard German because I thought she would otherwise not understand me in dialect, so my reaction was normal" (Lili, 01:35). The fact that Lili references this reaction as "normal" implies, however, that, in the words of Casasanto et al. (2010, p. 131), diverse "social and interactional factors may combine with automatic factors to produce linguistic accommodation," implying a sort of overlearned social accommodative behavior at play. Leo (01:29) further underscores this assumption: "[I]f someone doesn't [speak Austrian dialect], then you switch [to standard German] for better understandability, so more as a consideration for the counterpart […]. But that happens relatively automatically."

In sum, the sample pool shows an overall trend of convergence towards the variety of the respective interlocutor. While the SDs are rather large and thus demonstrate high variability between speakers, the participants' individual (differential) varietal behavior is a natural phenomenon for the Bavarian-Austrian context (e.g., Kaiser & Ender, 2021). Even given this fact, the between-participant trends of convergence (in particular toward the standard German-speaking interlocutor) speak greatly in favor of the VR discourse completion tasks' effectiveness in eliciting accommodation behavior with regard to the dialect and standard varieties. Whether such convergence mechanisms are of social-interactional nature or a result of interactive alignment processes is not entirely clear, though it does appear to be a certain mixture of the two. Due to the experiment's design, it was not possible for participants to achieve immediate social goals with the interlocutors. Yet, processes of convergence to the standard German-speaking interlocutor were still attributed to "ensuring comprehension," despite the fact that neither interlocutor could acknowledge any such social obligingness. This suggests a complex interplay of overlearned, internalized social factors in combination with mechanistic, automatic accommodation/convergence processes. In this vein, it can be argued that VR utilizing pre-filmed videos of human interlocutors are ecologically effective means of eliciting overlearned social (i.e.,

convergence to a non-dialect-speaking interlocutor for reasons of comprehension) and automatic convergence tendencies.

DISCUSSION

The overarching goal of the present contribution was to evaluate the use of a novel virtual reality elicitation methodology as a viable alternative to semi-directed/sociolinguistic interviews. This VR instrument is advantageous in several regards: the design allows researchers to maintain a high degree of experimental control over (a) stimulus choice and design, (b) power and interactional asymmetries between dyads, and (c) the degree of formality. The design moreover eliminates the possibility of unintentional interviewer accommodation to the interviewee and aids in reducing the artificial spatial divide between participant and stimulus classic to more traditional experimental procedures.

The study began by using qualitative data to shed light on the immersive, naturalistic, and communicative setting offered by the virtual reality instrument, which was paramount in determining VR's effectiveness in affording an ecologically valid elicitation environment. Overall, participants reported suppressing the objective reality to a large extent and foregrounded their high affective engagement in the virtual conversations, even if they did not experience a lack of total spatial dissociation. Such results underscored that the VR instrument presented a highly immersive, naturalistic, and communicative setting, emphasizing the ecological validity both of the instrument and the data elicited therewith.

In L2 sociolinguistic acquisition research, linguistic convergence has been used as an operational measure of sociolinguistic competence (Ender, 2019). The VR instrument introduced in this study was developed to elicit mechanisms of convergence from L2 learners to operationalize sociolinguistic competence. Once the high ecological validity of the VR instrument had been established, the second step was to evaluate VR's capacity to elicit mechanisms of linguistic convergence towards a respective interlocutor's language variety. In so doing, descriptive-quantitative data on participants' group-level convergence patterns were analyzed (see Supplementary Material for a Bayesian analysis) and the results linked to the observed patterns of the qualitative, introspective data on participants' rationale behind their conscious changes in varietal behavior. The speech data gave clear indications of between-participant tendencies to converge towards the respective variety of the interlocutor. In the retrospective interviews, participants further underscored that the rationale behind such varietal behavior was, indeed, that of linguistic convergence. Moreover, the participants referenced an interplay of social-interactional and mechanistic, automatic factors responsible for such convergence trends. It can be argued that this integrative approach, i.e., merging qualitative introspective data as an explanatory force with the descriptive-quantitative analysis as a means to buttress the weaknesses of each individual approach, provides strong evidence in favor of this VR instrument's effectiveness in eliciting speakers' overlearned and internalized mechanisms of linguistic convergence.

While there certainly exist other methods with similar goals of collecting sociolinguistic data, such as the (Labovian) sociolinguistic/semi-directed interview, VR offers an improvement due to its highly immersive nature and the experimental control that can be ensured. It can thus be argued that this elicitation instrument configuration and variations/adaptions thereof can capture speech data of more realistic quality, in that they are likely less prone to bias and distortion due to unperceivable yet very present social and situational asymmetries. Moreover, given the design at hand, unintentional accommodation by the interlocutors to the language of the participant can be entirely excluded. As a further advantage, by generating VR environments across myriad situations, settings and with different distributions of social distance, it is also feasible to assume that VR can elicit a large range of speech styles speakers encounter and use in naturally occurring speech settings, a

well-known shortcoming of more traditional sociolinguistic/semi-directed interviews (Schilling-Estes, 2002, 2008). In capitalizing on such aforementioned advantages, future studies making use of this or similar VR-based elicitation instruments can collect sociolinguistic data from L2 learners and operationalize sociolinguistic competence as, e.g., convergence to a VR interlocutor. This would allow designs similar to, e.g., Ender (2019) and Dewaele (2004a), whose operational measure of sociolinguistic competence was convergence, to exclude confounding factors (power asymmetries, unintentional interviewer accommodation) and so evaluative learners' capacity to engage in linguistic convergence in a more controlled experimental setting, without sacrificing ecological validity.

On a final note, it is a drawback of the present contribution that only native speakers were recruited, as the instrument is intended for use with L2 speakers. However, given L2 learners' diverse sociolinguistic repertoires, they were not the ideal subject pool for a pilot study evaluating the effectiveness of VR's capacity to elicit linguistic convergence. While research on how VR impacts L1 versus L2 speakers' performance is scarce, e.g., Taguchi (2022b) found that in a VR task focusing on speech rate of speech acts in high versus low PDR (power, distance, degree of imposition) situations, both L1 and L2 speakers spoke more slowly in PDR-high than in PDR-low situations. Moreover, both L1 and L2 speakers showed similar patterns of using more direct strategies and less modifications in PDR-low than PDR-high situations in the VR-based task. In Taguchi (2022a), conversely, L1 and L2 speakers behaved differently in terms of intonation in dispreferred speech acts, such that L1 speakers used more falling and L2 learners more rising tones. This phenomenon, however, may be attributed to L2 learners' limited experience with dispreferred acts in university settings, rather than to the VR instrument itself. With Taguchi's results in mind, it can be argued that, if L2 learners are capable of engaging in linguistic convergence to different varieties, this VR instrument can plausibly elicit such behavior, as it has done with the L1 speakers. Of course, this remains an avenue for future research, as task format might nonetheless differently influence L1 and L2 speakers.

CONCLUSION

The present contribution evaluated the use of virtual reality as an innovative elicitation methodology to compensate for several long-standing weakness of sociolinguistic/semi-directed interviews. The results of the present pilot study suggest that VR (a) presents an immersive, naturalistic, and communicative setting, underscoring its ecological validity as an experimental instrument, and (b) elicits overlearned behaviors/mechanisms of linguistic convergence, which can be used as an operational measure of L2 sociolinguistic competence. In using VR discourse completion tasks to elicit speech data, it is possible to control for otherwise confounding variables such as fluctuating social distance/dominance, unintentional interviewer accommodation to the interviewee and power asymmetries evident in (Labovian) semi-directed interviews, which have hitherto dominated the research designs addressing questions on L2 sociolinguistic acquisition. In further research on (particularly L2, but also L1) sociolinguistic acquisition, attempts need to be made to more rigorously constrain such aforementioned confounding variables to most reliably and effectively explore the processes mediating the development of sociolinguistic competence. The VR discourse completion tasks discussed in this study provide but one possibility to do so.

REFERENCES

Amin, A., Gromala, D., Tong, X., & Shaw, C. (2016). Immersion in Cardboard VR compared to a traditional head-mounted display. In S. Lackey & R. Shumaker (Eds.), *Virtual, augmented and mixed reality* (pp. 269–276). Springer.

Bardovi-Harlig, K., & Hartford, B. S. (1993). Refining the DCTs: Comparing open questionnaires and dialogue completion tests. In L. Bouton & Y. Kachru (Eds.), *Pragmatics and language learning, Selected papers presented at the Annual Meeting of the International Conference on Pragmatics and Language Learning* (4th ed., pp. 143–165). University of Illinois at Urbana-Champaign.

Billmyer, K., & Varghese, M. (2000). Investigating instrument-based pragmatic variability: Effects of enhancing discourse completion tests. *Applied Linguistics*, *21*(1), 517–552. https://doi.org/10.1093/applin/21.4.517

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. http://doi.org/10.1191/1478088706qp063oa

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

Coupland, N. (1985). 'Hark, hark, the lark': Social motivations for phonological style-shifting. *Language & Communication*, *5*(3), 153–171. http://doi.org/10.1016/0271-5309(85)90007-2

de Fina, A., & Perrino, S. (2011). Introduction: Interviews vs. 'natural' contexts: A false dilemma. *Language in Society*, *40*(1), 1–11. http://doi.org/10.1017/S0047404510000849

Dewaele, J.-M. (2002). Using socio-stylistic variants in advanced French interlanguage: The case of 'nous'/'on." *Eurosla Yearbook*, *2*(1), 205–226.

Dewaele, J.-M. (2004a). Retention or omission of the *ne* in advanced French interlanguage: The variable effect of extralinguistic factors. *Journal of Sociolinguistics*, *8*(3), 433–450. http://doi.org/10.1111/j.1467-9841.2004.00268.x

Dewaele, J.-M. (2004b). *Vous* or *tu*? Native and non-native speakers of French on a sociolinguistic tightrope. *International Review of Applied Linguistics*, *42*(2), 383–402. http://doi.org/10.1515/iral.2004.42.4.383

Drljača Margić, B. (2017). Communication courtesy or condescension? Linguistic accommodation of native to non-native speakers of English. *Journal of English as a Lingua Franca*, *6*(1), 29–55. http://doi.org/10.1515/jelf-2017-0006

ELAN. (2021). *ELAN*. Max Planck Institute for Psycholinguistics, The Language Archive. https://archive.mpi.nl/tla/elan

Ender, A. (2019). *Dialekt-Standard-Variation im ungesteuerten Zweitspracherwerb des Deutschen..Eine soziolinguistische Analyse zum Erwerb von Variation bei erwachsenen Lernenden* [Dialect standard variation in uncontrolled second language acquisition of German. A sociolinguistic analysis of the acquisition of variation in adult learners]. [Unpublished Habilitation thesis]. University of Freiburg.

Ender, A., & Kaiser, I. (2009). Zum Stellenwert von Dialekt und Standard im österreichischen und Schweizer Alltag – Ergebnisse einer Umfrage [On the importance of dialect and standard in everyday life in Austria and Switzerland - results of a survey]. *Zeitschrift Für Germanistische Linguistik*, *37*(2), 266–295.

Geeslin, K. L. (2010). Beyond "naturalistic": On the role of task characteristics and the importance of multiple elicitation methods. *Studies in Hispanic and Lusophone Linguistics*, *3*(2), 501–520. https://doi.org/10.1515/shll-2010-1084

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, N. Coupland, & J. Coupland (Eds.), *Contexts of accommodation. Developments in applied linguistics.* (pp. 1–68). Cambridge University Press.

Heyselaar, E., Hagoort, P., & Segaert, K. (2017). In dialogue with an avatar, language behavior is identical to dialogue with a human partner. *Behavior Research Methods*, *49*(1), 2351–2356. http://doi.org/10.3758/s13428-015-0688-7

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14–26. http://doi.org/10.3102/0013189X033007014

Kaiser, I. (2021). Children's linguistic repertoires across dialect and standard speech: Mirroring input or co-constructing sociolinguistic identities? *Language Learning and Development*, *18*(1), 1–21. https://doi.org/10.1080/15475441.2021.1922282

Kaiser, I., & Ender, A. (2021). Intra-individual variation in adults and children: Measuring and conceptualizing individual dialect–standard repertoires. *Linguistics Vanguard*, *7*(s2), 1–11. http://doi.org/10.1515/lingvan-2020-0032

Kleiner, S., & Knöbl, R. (2015). Einleitung [Introduction]. In Duden Verlag (Ed.), *Duden. Das Aussprachewörterbuch* (7. Auflage, p. 9–160). Dudenverlag.

Labov, W. (1966). *The social stratification of English in New York City.* Center for Applied Linguistics.

Labov, W. (1972a). *Sociolinguistic patterns*. University of Pennsylvania Press.

Labov, W. (1972b). Some principles of linguistic methodology. *Language in Society*, *1*(1), 97–120. http://doi.org/10.1017/S0047404500006576

Meyerhoff, M., & Schleef, E. (2012). Variation, contact and social indexicality in the acquisition of (ing) by teenage migrants. *Journal of Sociolinguistics*, *16*(3), 398–416. http://doi.org/10.1111/j.1467-9841.2012.00535.x

Milroy, L., & Gordon, M. (2003). *Sociolinguistics: Method and interpretation.* Blackwell.

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103. http://doi.org/10.1111/0022-4537.00153

Papachristos, N. M., Vrellis, I., & Mikropoulos, T. A. (2017). A comparison between Oculus Rift and a low-cost smartphone VR Headset: Immersive user experience and learning. *IEEE 17th International Conference on Advanced Learning Technologies* (ICALT), 477–481. http://doi.org/10.1109/ICALT.2017.145

Peeters, D. (2019). Virtual reality: A game-changing method for the language sciences. Psychonomic Bulletin & Review, 26(3), 894-900 https://doi.org/10.3758/s13423-019-01571-3

Peeters, D., & Dijkstra, T. (2017). Sustained inhibition of the native language in bilingual language production: A virtual reality approach. *Bilingualism: Language and Cognition*, *6*(1), 1–27. http://doi.org/10.1017/S1366728917000396

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

Regan, V. (1996). Variation in French interlanguage: A longitudinal study of sociolinguistic competence. In R. Bayley & D. Preston (Eds.), *Second language acquisition and linguistic variation* (pp. 177–201). John Benjamins.

Regan, V. (2010). Sociolinguistic competence, variation patterns and identity construction in L2 and multilingual speakers. *EUROSLA Yearbook*, *10*(1), 21–37.

Ritchie, T. D., Skowronski, J. J., Wood, S. E., Walker, W. R., Vogl, R. J., & Gibbons, J. A. (2006). Event self-importance, event rehearsal, and the fading affect bias in

autobiographical memory. *Self and Identity*, *5*(1), 172–195. http://doi.org/10.1080/15298860600591222

Schegloff, E. (1996). Turn-organization: One intersection of grammar and interaction. In E. Schegloff, E. Ochs, & S. Thompson (Eds.), *Interaction and Grammar* (pp. 52–133). Cambridge University Press.

Schilling-Estes, N. (2002). Investigating stylistic variation. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 375–401). Blackwell.

Schilling-Estes, N. (2008). Stylistic variation and the sociolinguistic interview: A reconsideration. In R. Monroy & A. & Sánchez (Eds). *25 años de Lingüística Aplicada en España: Hitos y retos* [25 Years of Applied Linguistics in Spain: Milestones and challenges]. Murcia: Editum, 971–986

Selting, M. (2000). The construction of units in conversational talk. *Language in Society*, *29*(4), 477–517. http://doi.org/10.1017/S0047404500004012

Stoyanchev, S., & Stent, A. (2009). Lexical and syntactic priming and their impact in deployed spoken dialog systems. *Proceedings of NAACL HLT 2009: Short Papers*, 189–192.

Tagliamonte, S. (2006). *Analysing sociolinguistic variation*. Cambridge University Press.

Taguchi, N. (2022a). Dispreferred speech acts in virtual reality: Analysis of tone choices and hesitations. *System*, *106*, 102793. http://doi.org/10.1016/j.system.2022.102793

Taguchi, N. (2022b). Immersive virtual reality for pragmatics task development. *TESOL Quarterly*, *56*(1), 308–335. http://doi.org/10.1002/tesq.3070

Vanrell, M. del M., Feldhausen, I., & Astruc, L. (2018). The discourse completion task in Romance prosody research: Status quo and outlook. In I. Feldhausen, J. Fliessbach, & M. del M. Vanrell (Eds.), *Methods in prosody: A Romance language perspective,* (pp. 191–227). Language Science Press.

Vergeiner, P. C. (2020). Sprachnormbezogene Akkommodation in der Hochschullehre [Language norm-related accommodation in university teaching]. *Zeitschrift Für Angewandte Linguistik*, *72*(1), 33–62. http://doi.org/10.1515/zfal-2020-2027

Wolfson, N. (1976). Speech events and natural speech: Some implications for sociolinguistic methodology. *Language in Society*, *5*(1), 189–209. http://doi.org/10.1017/S0047404500007028

SUPPLEMENTARY MATERIAL

USER EXPERIENCE QUESTIONNAIRE

RQ: To what extent do the virtual reality discourse completion tasks provide an overall positive user experience in terms of spatial presence, usability, simulator sickness, workload, and overall satisfaction?

This RQ refers to the overall user experience metric so as to evaluate VR in and of itself as a viable elicitation instrument. At the most basic level, we are interested in the participant-subjective degree of immersion provided by the instrument as well as its overall usability in terms of the five indices listed.

METHOD

Participants

User experience data from 12 participants are analyzed. The sample size differs slightly from the main body of the article as three participants' speech data were not analyzed: Two participants were excluded from the speech data analysis, as they had (while limited) previous knowledge of the parameters of the study (i.e. about the main focus on convergence tendencies) and one because his native language was not German, despite demonstrating native-like mastery.

Instrument

User experience questionnaire. The user experience questionnaire is adapted from Papachristos et al. (2017) and addresses the question of how well the virtual reality headset provides an immersive environment in terms of spatial presence, usability, workload, virtual reality sickness and satisfaction:

- Spatial presence refers to the 'immersiveness' of new technology and targets factors such as field of view and visual quality. The "enhanced sense of presence is central to the use, and therefore the usefulness and profitability, of the new technologies" (Lombard & Ditton, 1997) and thus constitutes an important measurement metric in the evaluation process. It was measured using an abridged version of the Temple Presence Inventory (Lombard et al., 2009), which comprises four items on 7-point scales.
- In measuring how easy the instrument is to use, the metric usability was included, as measured by the System Usability Scale (Brooke, 1996). This is a ten-item dimension with a range from 0 to 100. Scores above ~70 represent adequate usability.
- Virtual reality sickness was measured using the Virtual Reality Sickness Questionnaire (Kim et al., 2018), a nine-item dimension containing questions pertaining to oculomotor and disorientation symptoms. Scores can range between 0 and 100, with 0 representing 'no symptoms'. While there seems to be no agreed upon cutoff for 'acceptable', scores under 25 represent few to no symptoms.
- The metric workload, measured via the Task Load Index (Hart, 2006), was included to assess the physical and mental effort required in the VR task. This dimension contains six items (performance, mental demand, physical demand, temporal demand, effort, frustration) and scores between 0 and 100, with lower scores indicating a lower workload index.

- Following Papachristos et al. (2017), participants also reported their overall satisfaction with the virtual reality experience, a single-scale item on a seven point scale.

RESULTS

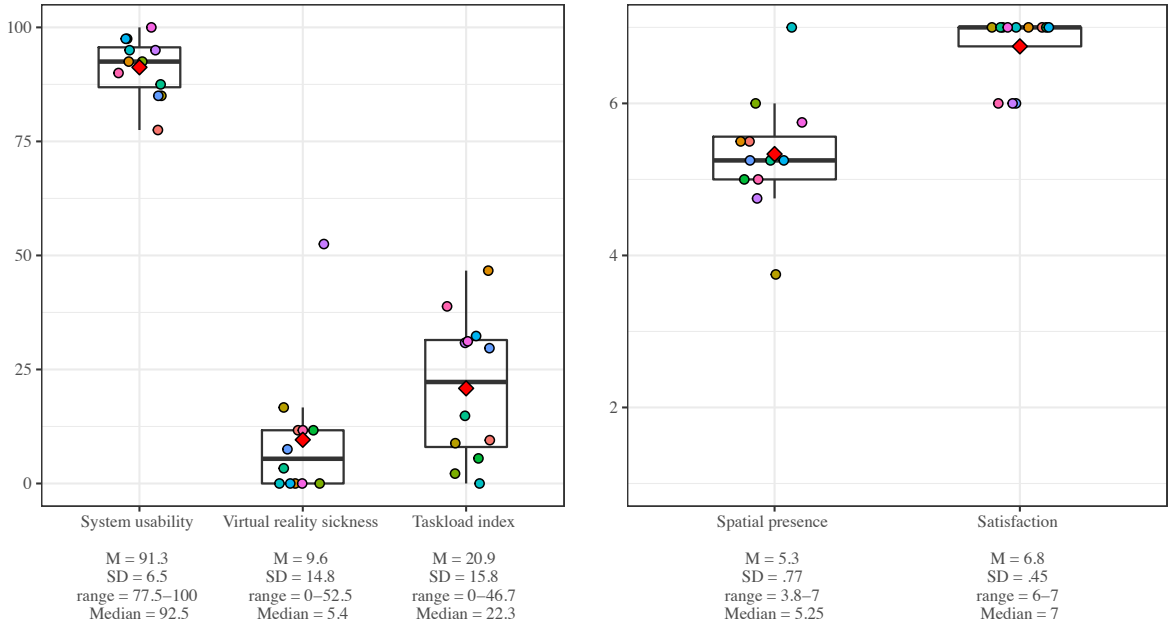
RQ: On the User Experience in VR

The virtual reality headset appears to provide an overall positive user experience in terms of spatial presence, usability, simulator sickness, workload and satisfaction (see SF 1). The following observations can be made based on the results:

- The spatial presence scores are relatively moderate. Given the 'immersive' nature of the virtual reality, higher scores were expected. However, even with more extensive VR configurations (e.g. Oculus), the spatial presence metric appear to remain within this vicinity (Papachristos et al., 2017, p. 479). Such moderate scores in this experimental setting could be attributed to the fact that participants were not able to navigate the environment freely, but were rather 'ushered' from interlocutor to interlocutor.
- System usability scores were found to be high in most cases, in any case higher than the ~70 threshold to represent 'adequate' usability. In fact, on the "Adjective Rating Scale" developed in Bangor et al. (2009, p. 119–121), the mean score (M = 91.3) falls into the category "best imaginable".
- The results for the taskload index indicate a moderate score, with the item 'mental demand' contributing the most to the overall workload, followed by 'effort' (see online supplementary information, supplementary figure 1). This could, however, simply be attributed to the fact that the participants were asked to answer a range of questions spontaneously. Participants otherwise conveyed feeling highly confident in completing the tasks given in the VR and further reported low scores of physical and temporal demand and low scores of frustration.
- The virtual reality sickness metric indicates that participants experienced few to no symptoms. Symptoms reported most often were eyestrain, difficulty focusing, blurred vision, and, to a lesser extent, fatigue, though the mean for each of these was well below 1 on a scale from 0–3 (see online supplementary figure 2) before weights and data transformation were applied for the overall VR sickness score.
- Finally, participants reported high levels of satisfaction with the virtual reality on the single-item scale.

SF 1

Results of the user experience questionnaire



Note. Each color represents a participant. The boxplots indicate the median and respective quartiles and the rhombus at the center represents the mean of the respective user experience metric. The summary statistics are outlined below each respective plot for the reader's convenience. 'System usability' 'Virtual reality sickness' and 'Taskload index' were measured on a 100-point scale, 'Spatial presence' and 'Satisfaction' on a seven-point scale; the metrics were thus visualized on separate plots according to their final scale.

Despite the 'cheaper' VR headset configuration, none of the domains display below-average scores, underscoring the immersive setting this technology can offer. While previous work has shown that cheaper VR headsets require higher physical demand than e.g. the Oculus (Papachristos et al., 2017, p. 480–481), the headset configuration used in this study did not appear to require a high amount of physical demand. Rather, mental demand and effort were reported to be the highest (albeit still far below the threshold for concern). In sum, virtual reality appears to provide an overall positive, spatially immersive user experience in terms of the metrics observed. The degree of 'immersiveness' will be discussed more extensively in what follows.

BAYESIAN ANALYSIS OF VARIETAL BEHAVIOR

RQ: Are there group-level trends of linguistic convergence to the respective varietal of the dialect-speaking and standard German-speaking virtual interlocutor?

I address this RQ in a hypothesis-testing fashion using a Bayesian mixed effects model, using which three hypotheses are tested:

H1. Dialect TCUs are produced on average more in interaction with the dialect-speaking interlocutor than with the standard German-speaking interlocutor.
H2. Mixture TCUs are produced on average more in interaction with the dialect-speaking interlocutor than with the standard German-speaking interlocutor.
H3. Standard TCUs are produced on average more in interaction with the standard German-speaking interlocutor than with the dialect-speaking interlocutor.

Data Analysis

Quantitative analysis. The statistical analysis entailed fitting a Bayesian multilevel regression model for factorial designs, using the brms package (Bürkner, 2017) in R (R Core Team, 2020).

I analyzed TCU count as a function of the within-subject fixed effects *interlocutor* (reference level 'dialect interlocutor') and *variety* (i.e. variety of the TCUs produced; reference level 'dialect TCU') and their two-way interaction. Given the outcome variable consists of count data and includes excess zeros, I used a zero-inflated negative binomial distribution, which maps the model estimates to the log space using the log linking function. The random effects specifications included the maximal random-effects structure justified by the design. The final model specifications are as follows:

formula = TCU ~ interlocutor * variety +
(1 | participant) + (1 | VRSet)
(0 + interlocutor * variety | participant)

The model estimates how individual participants vary in their overall TCU count and how the TCU count varies over the two VR sets (random intercepts). By-participant random slopes estimate how much TCU counts across participants vary according to the interlocutor and the language variety of the TCU, as well as their interaction. The Bayesian framework is moreover particularly suitable for obtaining reliable estimates when dealing with small samples, provided the prior is "set in the vague vicinity of the population value, even with a fairly large variance" (McNeish, 2016, p. 765). The model thus included regularizing, weakly informative priors, which were normally distributed and centered at 0 with a standard deviation of 2 (*Normal(0, 2)*). The model was fit with 8000 iterations (4000 warm-up) and Hamiltonian Monte-Carlo sampling was carried out with 8 chains. This yielded a posterior distribution of 32000 post-warmup samples for each tuple of parameter values.

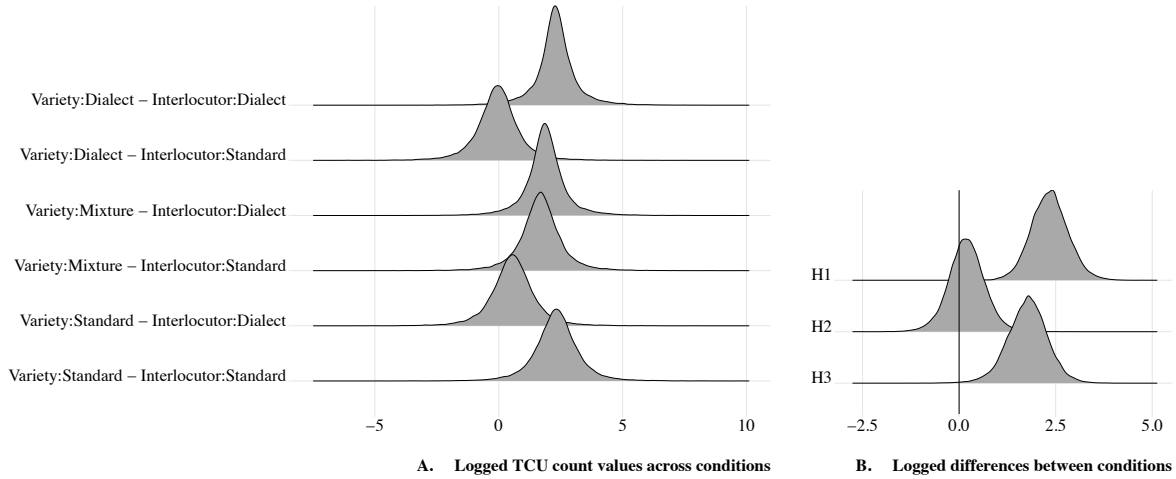
I make use of Bayesian hypothesis testing to determine how probable each of the aforementioned three hypotheses are by computing the probability from the posterior distribution that the difference δ between two conditions is larger than zero. I judge there to be compelling evidence for a given hypothesis if: (a) $\delta > 0$; (b) zero is not included in the 95% CI of δ and (c) the posterior $P(\delta > 0)$ is close to one (i.e. $\geq .95$, Franke and Roettger [2019]). See the online supplementary material for more information on Bayesian data analysis.

RESULTS

As a means to collect estimates of potential differences between the use of the three varieties with the two interlocutors, a Bayesian mixed-effects model was run (for a visual overview of the model summary, see online supplementary figure 5). By using the posterior distribution of this model, we can quantify our uncertainty around the respective estimates of the aforementioned potential differences corresponding to the three hypotheses to be tested. In fact, instead of settling on a binary accept/reject decision for each respective hypothesis, Bayesian inference allows us to determine the strength of evidence that a given hypothesis holds true, conditional on the data, model and prior assumptions. Bayesian data analysis is moreover particularly suitable when dealing with small sample sizes (see Data Analysis).

SF 2

Posterior density plots across interactional situations (A) and tested hypotheses (B)



Note. The first ridge plot (A) in Figure 2 approximates the posterior density for the logged estimates of the TCU counts in the individual varieties across the VR interactional conditions. For example, the first two distributions visualize clearly that dialect TCUs were produced to a far greater extent in interaction with the dialect-speaking interlocutor than with the standard German interlocutor. The second plot (B) displays the posterior distributions of the tested hypotheses, which assess whether the difference δ between two conditions (i.e. for H1 the difference between the amount of dialect TCUs in interaction with the dialect versus the standard German interlocutor) is larger than zero. The ridge plots are displayed on the log scale so as to be able to show the entire spread of values.

For easier interpretation, the point estimates of the posterior distributions displayed on the log scale in plots A and B in Figure 2 were exponentiated in order to report the following results on the original count scale (i.e. TCU count). As a whole, the results from the model lend weight to hypothesis 1, in that dialectal speech appears to be used more in interaction with the dialect interlocutor than with the standard German speaker. In interaction with dialect speakers, participants were expected to produce on average approximately 10 dialect TCUs (credible interval (CI) = [1.52, 65.0]), whereas in interaction with the standard German interlocutor, participants were predicted to produce only about 1 dialect TCU (CI = [0.13, 6.58]; see the first two ridge plots in Figure 2A). The results indicate that the raw difference between the amount of dialect TCUs produced with the dialect-speaking versus the standard-speaking interlocutor is approximately 10.4 (CI = [4.10, 27.5]; see H1 in Figure 2B). The posterior probability that this difference is larger than zero ($P(\delta > 0)$) is 1, i.e. 100%, giving us compelling evidence to conclude that the data and the model support hypothesis 1.

We find similar results for hypothesis 3. In interaction with the standard German-speaking interlocutor, participants were expected to produce on average roughly 10.3 standard TCUs (CI = [1.36, 78.5]), and with the dialect interlocutor only approximately 1.76 standard TCUs (CI = [0.23, 12.8]; see the last two ridge plots in Figure 2A). The raw difference between the amount of standard TCUs produced with the standard German interlocutor versus with the dialect interlocutor is roughly 5.85 (CI = [1.98, 16.0], $P(\delta > 0) = 1$; see H3 in Figure 2B). Despite compelling evidence for these two hypotheses, the credible intervals do cover a relatively large range of values, indicating a healthy amount of uncertainty in terms of the respective differences—a larger sample pool and/or more data from the speakers would be necessary to shrink the credible intervals and constrain our uncertainty of the respective differences.

Moreover, we must attribute even more uncertainty to hypothesis 2: In interaction with the dialect-speaking interlocutor, participants were expected to produce on average approximately 6.54 mixture TCUs (CI = [0.99, 44]), and with the standard German interlocutor 5.38 mixture TCUs (CI = [0.8, 40.7]; see the middle two ridge plots in Figure 2A). The raw difference in the amount of mixture TCUs produced with the dialect-speaking versus standard German-speaking interlocutor is roughly 1.22 (CI = [0.47, 3.15], $P(\delta > 0) = 0.67$; see H2 in Figure 2B), which underscores that there is a healthy amount of uncertainty associated with hypothesis 2.

REFERENCES

Bangor, A., Kortum, P., & Miller, J. (2009). Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. *Journal of Usability Studies*, *4*(3), 114–123.

Brooke, J. (1996). SUS - A quick and dirty usability scale. In P. W. Jordan, B. Thomas, I. L. McClelland, & B. Weerdmeester (Eds.), *Usability Evaluation in Industry* (pp. 189–194). Taylor & Francis.

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Franke, M., & Roettger, T. (2019). *Bayesian regression modeling (for factorial designs): A tutorial*. https://github.com/michael-franke/bayes_mixed_regression_tutorial/blob/master/text/bmr_tutorial.tex

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908.

Kim, H. K., Park, J., Choi, Y., & Choe, M. (2018). Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. *Applied Ergonomics*, *69*(1), 66–73.

Lombard, M., & Ditton, T. (1997). At the Heart of It All: The Concept of Presence. *Journal of Computer-Mediated Communication*, *3*(2), available under https://onlinelibrary.wiley.com/doi/full/10.1111/j.1083-6101.1997.tb00072.x.

Lombard, M., Ditton, T. B., & Weinstein, L. (2009). Measuring Presence: The Temple Presence Inventory. *Twelfth International Workshop on Presence*, 1–15.

McNeish, D. (2016). On Using Bayesian Methods to Address Small Sample Problems. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(5), 750–773.

R Core Team. (2020). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. https://www.R-project.org/