

Methodology and Research Practice

Using the Electronically Activated Recorder (EAR) to Capture the Day-to-Day Linguistic Experiences of Young Adults

Alessandra Macbeth^{1 a}, Michelle R. Bruni², Brittenny De La Cruz², Jacqueline A. Erens³, Natsuki Atagi⁴, Megan L. Robbins², Christine Chiarello², Jessica L. Montag^{3 b}

¹ Psychology, Azusa Pacific University, Azusa, CA, US, ² Psychology, University of California, Riverside, CA, US, ³ Psychology, University of Illinois, Urbana-Champaign, Champaign, IL, US, ⁴ Child and Adolescent Studies, California State University, Fullerton, CA, US

Keywords: corpus linguistics, speech production, language production, naturalistic language, electronically activated recorder

<https://doi.org/10.1525/collabra.36310>

Collabra: Psychology

Vol. 8, Issue 1, 2022

Both recently and historically, naturalistic datasets and corpus analyses have played an important role in the formulation and testing of key theories and hypotheses in language development and use. The present work details ways in which an existing tool, the Electronically Activated Recorder (EAR), can be used in the cognitive and language science domains to better understand the content of day-to-day speech. From our sample of 75 young adult college students – a population with diverse linguistic experiences – we found enormous variability in the total amount of speech produced and the number of unique words spoken. Further, we discovered that individuals who speak frequently may not be the same individuals that produce long utterances, and we quantified the contexts in which individuals tend to speak. We argue that studies examining naturalistic speech in adults are rare, and through our data, we aim to demonstrate how the EAR can be used in novel ways to create both individual and group-level corpora of adults' spoken language use.

1. Introduction

The field of psychology has made significant discoveries via the collection and analysis of naturalistic data. Such advances have made these observations of naturalistic environments more logistically and computationally possible than ever before. The data obtained from these observations have been valuable for gaining insight into both cognitive and developmental processes, as well as for theory development. In the present work, we discuss how the Electronically Activated Recorder (EAR), a time sampling methodology which has previously been used to collect audio samples of individuals' daily lives, can be used in linguistic contexts to address various questions about language use. Specifically, we focus on how the EAR is an ideal tool for understanding the diverse day-to-day linguistic experiences of young adults.

1.1. The Natural Environment and Individual Experiences

Gaining a better understanding of the natural environment and an individual's experiences have led to important practical and theoretical advances in the cognitive sciences. For example, head-mounted cameras and eye-trackers have provided insight into the sorts of visual experiences that infants encounter, with major implications for visual development, including face perception and social behaviors such as gaze following or gesture recognition (Fausey et al., 2016; Franchak et al., 2011; Slone et al., 2018). Furthermore, in the social domain, differences in the visual complexity of real city scenes may be an environmental influence that underlies cultural differences in holistic versus analytic scene perception (Nisbett & Miyamoto, 2005). In the field of motor development, children's levels of physical activity are associated with individual differences in various motor skills (Fisher et al., 2005), and in the clinical domain, collecting finer-grained information via experience sampling of individuals' mood, stressors, and interpersonal

a Please direct correspondence to Alessandra Macbeth, Department of Psychology, Azusa Pacific University, 901 E. Alosta Avenue, Azusa, CA, 91702
Email: amacbeth@apu.edu

b Please direct correspondence to Jessica L. Montag, Department of Psychology, University of Illinois, Urbana-Champaign, 603 E. Daniel Street, Champaign, IL, 61820
Email: jmontag@illinois.edu

interactions may enable clinicians to improve patient assessment and treatment (Myin-Germeys et al., 2018). By documenting and understanding the environment in which the individual is immersed, we can better understand how cognitive processes might be shaped by or interact with specific patterns in the individual's environment. Without this naturalistic data, certain mechanisms and explanations may elude us.

Many insights from naturalistic observation across multiple areas of psychology research come specifically from methodologies that employ the EAR, a time sampling method that records snippets of audio from a participant's day-to-day conversations at intermittent intervals (Mehl, 2017; Mehl et al., 2001). This method has advantages over other types of ecological assessment, such as experience sampling, because it does not require self-report, puts little burden on the participant to document their lived experiences, and also does not require the researcher to extensively train the participant about how to document such experiences (Mehl, 2017). The EAR is a passive and relatively unobtrusive modality of data collection. Participants are unaware of when the device is recording, which increases the likelihood that speech elicited by the wearer is natural and not contrived. Much of the research to date that has used the EAR has focused on topics of social, health, and personality psychology, including assessments of moral behavior (Bollich et al., 2016), personality assessment (Allemand & Mehl, 2017; Mehl et al., 2006) — including in patients with schizotypy (Minor et al., 2018) — and speech patterns of individuals coping with chronic illness or breast cancer (Karan et al., 2017; Robbins et al., 2011, 2018, 2019). In the cognitive arena, the EAR has been used largely to assess episodic and semantic detail in autobiographical memory (Wank et al., 2020), as well as the presence and linguistic complexity of speech in various social contexts (Demiray et al., 2020; Luo, Robbins, et al., 2019; Luo, Schneider, et al., 2019). Research using the EAR has provided remarkable insight into real-world patterns of verbal behavior, and has helped researchers relate these patterns of behavior to different psychological phenomena. There are clear external validity challenges associated with studying some social, health, and personality phenomena in laboratory settings, and obvious practical challenges associated with studying these phenomena outside of the lab. The EAR has provided a simple and powerful tool to measure naturalistic language behavior outside of the typical laboratory setting.

The power of gaining an understanding of the naturalistic environment is particularly pronounced in the field of language. Understanding characteristics of the language(s) to which children and adults are exposed is key for both the interpretation of behavioral data and theory development. However, the different aims of the fields have meant that child language learning researchers and adult language processing researchers have collected information about the natural language environment using different methodologies and for different goals. Both literatures serve as inspiration for the present work.

1.2. Naturalistic Language in Children

Currently and historically, there has been much interest in using features of children's naturalistic language environments to better understand language learning processes. There has been less emphasis in the adult literature on collecting and investigating patterns of spoken natural language, so much of the theoretical and practical motivation of the present work stems from the child literature.

For decades, researchers have been making audio recordings of young children: single recordings of a few minutes, longform recordings, or multiple short recordings over periods of weeks or months. Some investigations have focused on the child's own speech (e.g., Bloom et al., 1975; Brown, 1973; de Villiers & de Villiers, 1973; Vihman et al., 1985; among many others) to better understand developmental trajectories of productive language. Other investigations have focused on speech that is addressed to or available to the child (e.g., Brent & Siskind, 2001; Cartmill et al., 2013; Goldin-Meadow et al., 2014; Hart & Risley, 1995; Hirsh-Pasek et al., 2015; Hoff & Naigles, 2002; Hurtado et al., 2008; Huttenlocher et al., 1991; Ramírez-Esparza et al., 2017; Rowe, 2008; Snow, 1977; among many others). This work has been important for our understanding that the language environment is remarkably rich and structured, and plays a substantial role in driving language learning, with consequences for data interpretation and theory development (Lieven, 2016). The studies listed above often predict behavior, including individual differences in behavior in laboratory tasks, as a consequence of the linguistic and non-linguistic content of audio recordings.

Other lines of work combine the shorter recordings to create a large corpus (e.g., the CHILDES corpus; MacWhinney, 2000) that can predict normative trends in language learning trajectories of groups of children (Braginsky et al., 2019; Goodman et al., 2008; Hills et al., 2009, 2010; Swingley & Humphrey, 2018; Willits et al., 2014). These studies have meaningfully linked language input to language learning and have proposed environment-based explanations for laboratory phenomena. By understanding the day-to-day language that children encounter, at both an individual level and normative aggregations, we have gained a better understanding of the data that drives language learning.

In more recent work, researchers have used small, unobtrusive, audio recorders to record a full day or multiple days of a child's auditory environment (e.g., the Language Environmental Analysis (LENA) system) (Ford et al., 2008; Gilkerson & Richards, 2008). This data can answer different questions than audio that is recorded in a single setting (Bergelson et al., 2019; Casillas et al., 2020; Gilkerson et al., 2018; Mendoza & Fausey, 2021; Montag, 2020; Oller et al., 2019; Pretzer et al., 2019; VanDam et al., 2016). Much like the time sampling methodology employed in the present work, these day-long recordings allow for collection of data that can reveal temporal dynamics and contingencies that would otherwise not be possible.

The lessons learned from the use of naturalistic language data in child language development directly informs the present work. First, understanding the language that children produce or encounter—for both individuals and aggregating over individuals—has led to a better understand-

ing of the data on which language learning proceeds, which has consequences for theory development, data interpretation, and avenues for future research. Second, longform recordings capture day-long dynamics that provide unique information that shorter recordings do not. We take these lessons from the child literature as we explore ways to use naturalistic data from adults' language productions to gain insight into language processing and use.

1.3. Naturalistic Language in Adults

As in child language development, there is a long history of using corpora to predict adult behavior in various linguistic tasks. For example, statistics in text corpora predict not only group-level word naming and lexical decision latencies (Adelman et al., 2006; Balota et al., 2004; Bates et al., 2003; Brysbaert & New, 2009) but also various measures of semantic knowledge such as semantic priming, similarity ratings, or categorization (Huebner & Willits, 2018; Jones & Mewhort, 2007; Lund & Burgess, 1996; Olney et al., 2012; Pereira et al., 2016). Likewise, corpus analyses have proved to be helpful in predicting various measures of sentence processing (Garnsey et al., 1997; Gennari & MacDonald, 2009; Hare et al., 2007; Levy, 2008; Montag & MacDonald, 2015; Reali & Christiansen, 2007; Trueswell et al., 1993). Analyses of large corpora have given us enormous insight into the patterns that exist in typical language and into the knowledge that underlies skilled language use.

Despite many similarities, two key differences exist between the use of language corpora in the fields of child language learning and adult language processing. First, while naturalistic data from individual children's language environments might be used to predict individual differences or used in aggregate to predict normative behavior, the adult literature has primarily focused on the latter. While better theories for linking statistical properties of corpora to human behavior have certainly been a major focus of the field (Adelman et al., 2006; Zevin & Seidenberg, 2002), characteristics of the corpora themselves have been a major focus of research. One of the many themes that emerge from the adult literature is that prediction and explanation of behavior (such as those described in the previous paragraph) can be improved with a "better corpus." The underlying sentiment is that shortcomings in predicting behavior from corpora often derive from insufficient corpora rather than theoretical links between input and behavior. Thus, improving corpora is a means of better understanding human behavior. "Better corpora" can be described in a number of ways, including size (Burgess & Livesay, 1998; Recchia & Jones, 2009) or representativeness of the language contained for human experience (Brysbaert et al., 2011; Brysbaert & New, 2009). Some work has attempted to build more tailored corpora to predict individuals' behavior, as a consequence of unique linguistic experiences that an individual is likely to have (much like we observe in the child literature; Johns & Jamieson, 2018), but such endeavors represent newer trends in the field. Developing individuated corpora as a means of predicting differences across participants has not been as much of a focus of the adult literature as it has been in the child literature.

The second key difference between the adult and child literatures is that while the naturalistic corpora used to predict child language behaviors are largely *spoken*, the corpora used to predict adult language behaviors are generally *written*. For example, commonly used adult corpora include newspaper text (WSJ), textbooks (TASA; Touchstone Allied Science Association; <http://lsa.colorado.edu/spaces.html>), online materials (Wikipedia or Usenet), or movie subtitles (Subtlex; Brysbaert & New, 2009). Though movie subtitles reflect spoken dialogue, they are not spontaneous speech—rather, they are actors reciting written texts—so Subtlex may contain some similarities and differences to canonical written texts. This discrepancy in the written versus spoken domain likely derives both from the relative ease with which written corpora can be compiled relative to spoken corpora, and that for many literate adults, text is indeed an important source of language input. However, there are often profound differences between the language patterns contained in speech and text (Biber, 1988; Hayes, 1988; Montag & MacDonald, 2015; Roland et al., 2007). Moreover, differences in text exposure are known to meaningfully affect language experience, such that text exposure affects sentence comprehension and production behavior (Arnold et al., 2018; Cunningham & Stanovich, 1998; Montag & MacDonald, 2015; Payne et al., 2012; Street & Dabrowska, 2010). Perhaps one means toward developing better corpora is to better understand 1) differences between written and spoken language and 2) the extent to which adults' language experience may derive from each source.

The lessons learned from the adult literature complement those learned from the child literature. First, profound differences can exist in how corpora were built, such that some corpora make better predictions for data than others. Given that a corpus may fail to explain data due to either flawed theory or a flawed corpus, developing better corpora is important for data interpretation and theory development. The EAR may be able to help us build better corpora, by providing estimates of distributional properties of spoken language. Second, important differences exist between written and spoken language, and understanding the dimensions along which the domains vary may be practically and theoretically important for predicting, explaining, and theorizing about adults' language behavior.

1.4. Why Collect Naturalistic Language Data from Adults

The field of language research, broadly defined, may be able to make large theoretical advances surrounding the role of language experience in language use with the expanded collection of naturalistic language data from adults. First, studies with bilingual or multilingual participants generally rely on self-report to characterize participants and understand their language histories (Anderson et al., 2018; Li et al., 2020; Marian et al., 2007). To date, little is known about the degree to which self-report measures accurately reflect an individual's real language background and experiences. Participant samples that include bilingual and multilingual speakers, such as our sample, would allow researchers to better document if, when, how much, and with whom speakers use their multiple languages. Natu-

realistic data would allow researchers to have an additional measure of language experience in the form of documentation of the languages that the individual speaks day-to-day, potentially compared with or augmented by self-report.

Second, many aspects of spontaneous speech are difficult to study outside the laboratory. Spontaneous speech can be remarkably messy; by some estimates, adult speech contains about one error for every 1,000 words (Garnham et al., 1982). Spontaneous speech is also characterized by copious disfluencies, including pauses and fillers such as “um” or “uh” (Clark & Tree, 2002). Errors and disfluencies have been studied descriptively as a means of not only describing the language production system by understanding the systematicity in errors (Dell & Reich, 1981; Fromkin, 1973; MacKay, 1972), but also understanding individual differences in spoken language (Dell et al., 1997; Tausczik & Pennebaker, 2010). However, with the exception of some decades-old corpora (e.g., London-Lund corpus: Svartvik & Quirk, 1980; Switchboard corpus: Godfrey & Holliman, 1995), many speech error datasets were collected via laboratory tasks designed to elicit errors. The recording of naturalistic speech from adult participants would enable investigations of important language phenomena, including speech errors and disfluencies, that would otherwise be difficult to observe outside of laboratory tasks.

Third, as mentioned previously, many of the corpora used to describe or predict adult behavior derive from written texts rather than spontaneous spoken language. Additional data about the distributional statistics of spoken language may be warranted to build better estimates of the language that adults encounter day to day. Some language behaviors may be better explained by either spoken or written language such that the availability of both domains may improve predictive and explanatory power. Further, individuals may vary in both written and spoken language habits, and the lack of documentation about individual variability in spoken language habits leaves many potential questions of individual differences in language behavior unanswered.

1.5. The Present Study

The EAR has previously been used in many instances to collect data from college-aged samples (Manson & Robbins, 2017; Mehl et al., 2006; Mehl & Holleran, 2007; Mehl & Pennebaker, 2003). Our study aims to use the EAR in a novel way as a method to collect data that can answer questions that are specifically linguistic in nature. For example, the EAR can provide measures of language experience that might be used to predict laboratory-based language behavior or used in conjunction with self-report assessments of language habits. We believe this work represents a practical method by which individual differences in language use can be captured to make individual predictions or can be aggregated across multiple participants to build a corpus representative of college students' diverse spoken language experiences.

In addition to better understanding day-to-day spoken language use, our data collection will allow us to better understand spoken language habits of individuals who speak multiple languages. Our participant sample includes many speakers of other languages in addition to English, includ-

ing many heritage language speakers who learned a language at home either prior to or alongside English. Language habits of bilingual speakers are typically investigated via self-report measures, so this work may allow us to better understand the linguistic experiences of young bilingual and multilingual speakers, especially a group of bilingual speakers, heritage language speakers, whose experiences are relatively less represented in the literature.

The research reported here encompassed several aims. We first wanted to assess the overall presence of speech in our audio files, and where and with whom that speech occurred. Next, we measured the relative use of different languages, and assessed both absolute and relative amounts of speech produced by individual speakers. We then compared the audio collected on weekdays on weekends, to understand differences in overall language use by day, especially bilingual speakers' use of their multiple languages. Finally, we turned to the lexical content of the transcribed utterances. We mention the pitfalls of using lexical diversity as an individual difference measure of spoken language, and demonstrate how the lexical inventory gathered by the EAR compares to inventories of other major language corpora commonly used to measure word frequencies. Through these analyses, we demonstrate the utility of the EAR as an important tool - one that is relatively new to language researchers as well as others in the field of cognitive psychology - for measuring spoken speech patterns among diverse adult populations.

2. Method

2.1. Participants

Our sample was composed of undergraduate students from a large research university in Southern California. The study was advertised through the psychology department's participant pool. We collected data in two waves: 34 undergraduates participated in Wave 1, and 49 undergraduates participated in Wave 2. There were slight methodological changes made between the two waves of data collection, discussed in greater detail below. Seventy-five participants, 30 from Wave 1 and 45 from Wave 2, were included in the final dataset ($M_{\text{age}} = 19.21$ years, $SD = 1.41$; 50 female). Participants received \$25 and participant pool credit for their participation.

Reflecting the linguistic diversity of our Southern California sample, our participants spoke a variety of different languages. During Wave 1, we included participants with any language background, and in Wave 2, we included only participants who self-identified as bilingual. Nearly all participants (97.37%) reported some proficiency in a language other than English, and 76.32% spoke a non-English language in one or more of their valid audio files. Languages (and the number of participants who spoke each language) captured in the recordings other than English included: Amharic (1), Arabic (3), Burmese (1), Cantonese (1), Farsi (2), Hindi (1), Japanese (1), Korean (2), Mandarin (7), Portuguese (1), Punjabi (1), Russian (1), Spanish (33), Taiwanese (1), Teochew (1), Thai (1), and Vietnamese (6).



Figure 1. A) Widget image for the EAR app in the Google Play store. B) The interface that experimenters see after logging in to the EAR app (left), and what participants see once the recordings have begun (right). C) The EAR in its protective case and waist clip. D) The image printed on the bystander buttons and the stickers affixed to the phone carrying cases and armbands.

2.2. Materials & Procedure

The EAR is a free phone app which is compatible with Android devices. We downloaded the EAR from the Google Play store (see [Figure 1A](#)) onto Motorola Moto E (2nd Generation) phones. Using the EAR interface, we selected a recording duration and interval, recording start time, recording end time, and nightly six-hour blackout period (see [Figure 1B](#)). The EAR was programmed to record for 40 seconds every 12 minutes. This duration and interval were determined to be appropriate after pilot testing demonstrated that 40 seconds was long enough to capture several sentences of speech, permitting linguistic analyses of interest, and the 12-minute interval produced a reasonable number of audio files for analysis from each participant.

During the first laboratory session, participants were informed about the nature of the study, what types of sounds the EAR is designed to pick up (e.g., the participant's voices and sounds in the immediate environment, such as a TV or a car honking its horn), information about the recording duration and interval, and the safeguards in place to protect participant privacy. Participants were asked to wear the EAR as much as they were comfortable, including at home, school, and public places like a park or mall. The only context in which participants were explicitly told not to wear the EAR was at work, in order to avoid potential conflict with their supervisors or companies who might not want them participating in a research study while at work. Participants either wore the EAR from Thursday through Sunday or Friday through Monday, in order to capture any potential variations in speech production on weekdays versus week-

ends. Participants carried the EAR on their person either in a protective plastic case with a waist clip attached, or in an armband with a clear plastic covering.

Recordings began immediately after the participant's first laboratory session and ended when the participant went to bed on the fourth night or at midnight, whichever came later. These procedures documented above are consistent with the EAR best practices laid out in previous work (e.g., Kaplan et al., 2020; Mehl, 2017) and the EAR Repository scripts and guides maintained by Robbins and colleagues (2018) available at the Open Science Framework (<https://osf.io/n2ufd/>). Once the recording period was over, participants returned the EAR and completed a series of questionnaires related to their language background and experience with the EAR. Wave 2 participants also completed several behavioral tasks that will be documented more thoroughly in future work (e.g., Macbeth et al., 2022).

2.3. Ethical Considerations

Several safeguards were implemented to ensure the confidentiality and privacy of participants and their conversation partners, consistent with past EAR procedures (Kaplan et al., 2020; Robbins, 2017).

2.3.1. Participant privacy.

The EAR methodology includes two features to ensure that experimenters never hear conversations which the participant prefers to keep private. First, participants have the option of "pausing" the EAR at any time while wearing the device. Participants can open the EAR app and press the "Privacy" button on the home screen. This pauses the device for a set period of time (5 or 15 minutes, see below), and they can press the button as many times as they wish. Second, when participants returned the EAR, they were given the option to listen to their audio files and to identify any files that they wanted deleted. If the participant did want files deleted, the experimenter would delete them with the participant watching. These deletions were done before the researchers listened to any audio files, giving the participant autonomy over their recorded data.

2.3.2. Conversation partner privacy

California is a two-party consent state, meaning that all parties involved in a recorded conversation must consent to being recorded. To allow participants' conversation partners the ability to "opt out" of being recorded, participants wore carrying cases and buttons with the words, "This conversation may be recorded," and a picture of a microphone ([Figures 1C and 1D](#)). Participants were also asked to explicitly inform others that they interacted with about the possibility that their voices could be recorded (Manson & Robbins, 2017), so that potential conversation partners could choose whether to continue their conversation with the participant.

In addition, we coded only minimal information about conversation partner speech. The only information extracted from individual audio files into our language transcriptions was 1) a code indicating that a person other than the consented participant was speaking, 2) the language

in which the conversation partner was speaking, and 3) whether the conversation partner was male or female. No actual utterances of the conversation partners were transcribed.

2.4. Transcription, Coding, and Data Processing

Each participant's audio files were transcribed and coded by at least two different research assistants. Coders used Express Scribe transcription software, which is available for both Mac and PC devices (<https://www.nch.com.au/scribe/index.html>). Coders could start, stop, rewind, and fast-forward files via Express Scribe's computer interface or through a connected Infinity USB foot pedal. All transcriptions and codes were entered into a separate Excel sheet for each participant. Research assistants underwent extensive training (a minimum of two weeks, led by a member of the research team) before they transcribed and coded the audio files. A short compilation of "helpful hints" for transcribing and coding that we presented to coders as part of their training (e.g., how to code for pauses, speech in different languages, and unintelligible speech; transcribing slang, numbers, and translations) can be found in an online supplement at <https://osf.io/mpn4x/>. Coders had access to these materials at all times and could refer to them while transcribing and coding. After training, research assistants were required to transcribe and code a standard set of practice audio files before they began working with participant data.

2.4.1. Transcription

Table 1 includes examples of participant speech, to demonstrate some of the different speech patterns evident in our sample. Transcripts were typed verbatim, including partial words, disfluencies, or slang (e.g., "gotta," "cuz"). Importantly, in order to generate an accurate word count, these slang words were standardized and spelled consistently across coders. These preferred spellings are included in the online supplement mentioned above.

The speech of conversation partners was indicated by a series of three letters (e.g., xxx = English speech by the conversation partner), which helped maintain conversation structure. All participant English speech was transcribed. When the participants spoke a non-English language, their speech was transcribed literally by research assistants fluent in the target language and then translated into English. For some languages, we sought fluent speakers in other labs in the department to transcribe and translate our audio, to ensure that each recording was being transcribed by an individual fluent in that language. For only two languages (German, two files; Portuguese, six files) a participant's speech could not be transcribed and translated into English because we could not locate an individual familiar with the given language. The examples in Table 1 demonstrate audio files in which the entire 40-second recording interval is filled with speech (either by the participant or conversation partner), but files with a single word or phrase spoken by the participant (e.g., "okay," "yeah that's it") were also prevalent.

2.4.2. Coding and coder reliability

After the speech was transcribed, audio files were coded using the Social Environment Coding of Sound Inventory (Mehl & Pennebaker, 2003). Additionally, we coded the language(s) the participant and conversation partner(s) spoke, as well as technical aspects of each file (e.g., day of the week, sound quality problems, whether the participant discussed the EAR). For specific information about coding categories and intraclass correlation coefficients (ICC) across coders, see Appendix A.

2.5. Changes from Wave 1 to Wave 2

Based on our own observations and participant feedback, we made minor changes to the procedure between the first and second waves of data collection. First, we offered an armband option because certain clothing choices (e.g., dresses, skirts) prohibited participants from wearing the EAR on their waist. In Wave 2, six of 46 participants chose the armband option. Second, we adjusted the length of the privacy setting. During Wave 1, the privacy button paused the EAR for five minutes. Some participants pressed the button several times in a row and deleted files that were captured in between these short pauses. Therefore, we changed the privacy interval to 15 minutes to ensure participant comfort and privacy. Third, we discovered that it was sometimes difficult to discern the participant's voice from that of conversation partners. To address this challenge, we recorded a baseline speech sample for Wave 2 participants. Before leaving with the EAR, participants were asked three questions ("What do you like to do for fun?", "What did you do last weekend?", and "What are your plans after graduation?") and provided responses to these questions in English and in their most dominant non-English language. If transcribers had difficulty identifying the participant's voice in an audio file, they could refer back to the speech sample to determine which voice belonged to the participant.

Finally, we adjusted our coding and transcribing procedure slightly to make coding more efficient. In Wave 1, two coders independently listened to, transcribed, and coded a participant's entire set of audio files. In Wave 2, coding was done in two steps. In step 1, two coders independently transcribed all files with speech, and then only coded categories that denoted 1) whether there was a problem with the audio file, 2) if the participant was sleeping, 3) if the participant was speaking, and 4) the language that the participant or any conversation partners spoke. In step 2, two additional coders independently verified the preliminary codes of the first two coders, and then completed all additional codes only for the audio files with speech (participant or conversation partner). This way, the bulk of the file coding was done after the files that did not contain any meaningful linguistic information were identified, which was a more efficient workflow.

Table 1. Examples of the types of speech that can be extracted from participants' audio files.

Audio File Type	Original Transcript	Translation
English Only	I told you something I went back to sleep and you came for a second time xxx no because the dog was in the room she took a nap with me and she laid in the bed and I fell asleep and she fell asleep on the pillows next to me xxx oh I don't know she was laying down xxx yes and then when I woke she was at the door and she wanted to leave and I assumed you were here xxx huh xxx	
Code-Switch (English & Thai)	So we shall see, like I said nothing we can do about it now, just gotta move forward [โอ้โหมีรถไฟอีก] ttt [มันไม่ช่วยออกแต่มันจะเขียนไว้] like* it shows up but it it can't physically add on cuz you're not technically not undergrad anymore so this will come as like* your grad your postgrad but it still shows like* okay this kid got an A, you know?	So we shall see, like I said nothing we can do about it now, just gotta move forward [oh my there's a train] ttt [it won't help, but it will be written] like* it shows up but it it can't physically add on cuz you're not technically not undergrad anymore so this will come as like* your grad your postgrad but it still shows like* okay this kid got an A, you know?
Non-English Only (Spanish)	[nos quedamos] sss [mañana les puedo tomar unas fotos pa tenerlas ya listas] sss mhm sss [no na mas sábado y domingo] sss [en el verano ya vere si hago entre semana o lo que sea pero si ya va er pal verano] sss	[we will stay] sss [tomorrow I'll be able to take the photos so they're ready] sss mhm sss [no, only Saturday and Sunday] sss [in the summer I'll see if I do it in the week or whatever but that will be in the summer] sss

Note. Three underlined letters indicate a placeholder where the conversation partner is speaking; xxx = English, ttt = Thai, sss = Spanish. Non-English participant speech is enclosed in brackets in both the original transcript and the English translation. Use of the word "like" as a filler or disfluency is coded during transcription as "like*".

2.6. Coding Schemes and Natural Language Processing

Text analysis code was written in Python, and data analysis was performed in R. All code used for our analyses is included at <https://osf.io/mpn4x/>. For further recommendations for efficient coding and transcribing schemes that may aid subsequent computer code for data analysis, see Appendix B.

3. Results

We first describe participant compliance to gauge the extent to which participants wore the EAR. We then discuss analyses of the speech patterns captured by the EAR, as detailed in the research aims mentioned previously.

3.1. Participant Compliance

We first verified that participants wore the EAR as instructed to ensure we gathered a representative sample of the participant's day-to-day speech. Overall compliance with wearing the EAR was high. Participants were excluded due to noncompliance if there was virtually no intelligible speech in the entire set of audio files. It is very easy to tell when participants did not comply with proper wearing procedures, because there was either no ambient sound in the audio file or the sound was muffled in some way (e.g., the EAR was left in a backpack or purse). Seven participants were removed prior to analysis due to a failure to wear the EAR as instructed (four from Wave 1, three from Wave 2).

We assessed compliance in two ways to gain converging insight into participants' EAR wearing habits, consistent with past EAR protocols (Manson & Robbins, 2017; Mehl & Holleran, 2007). The first assessment was a self-report question given to participants at study completion: "Over the last four days, what percentage of the day (based on

your time awake) were you carrying the EAR immediately on you (0-100%)?" Participants reported wearing the EAR during 79.1% of their waking hours ($SD = 14.6\%$, range = 40-100%). Compliance was also assessed by calculating the proportion of files in which coders suspected the participant was not wearing the EAR, based on acoustic features of the recordings (e.g., there was no ambient sound or movement directly around the EAR). Using this calculation, participant compliance was 81.4% ($SD = 16.8\%$, range = 38.3-100%). Consistent with previous findings (Robbins et al., 2014), participants' self-report of their compliance and EAR-assessed compliance, were indeed moderately correlated, $r(74) = .53$, $p < .001$. Thus, we can conclude that participants regularly wore the EAR, and we collected at least a somewhat representative sample of participants' day-to-day speech.

3.2. Presence of Speech

We began by computing the presence/absence of speech ratio for each participant, which can help estimate how much language one might produce in a day and may be used to predict behavior in lab-based tasks assessing various aspects of speech production. On average, 300.1 audio files per person ($SD = 45.2$, median = 314, range = 91-337 files) were recorded. As is standard practice with EAR data, all audio files in which participants were sleeping were removed from further analysis, as were any files that participants chose to delete. Thirty participants (out of 76) chose to delete one or more audio files. One participant attempted to be helpful and deleted all of their audio files without speech, resulting in 231 deleted files. This participant was excluded, resulting in 75 total participants for all subsequent analyses. The remaining 29 participants who deleted audio files deleted an average of 6.9 files ($SD = 7.9$, median = 4, range = 1-36).

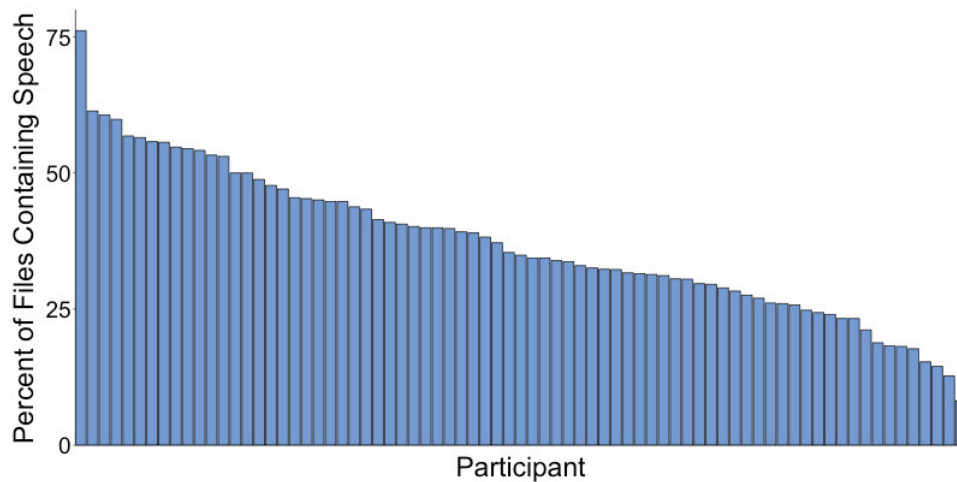


Figure 2. The percentage of valid audio files in which there was speech, for each of the 75 participants.

Next, we eliminated audio files that posed problems for interpretation, including those with 1) a zero-second long recording, suggesting a device/app malfunction, 2) poor recording quality (e.g., loud noises that drowned out participant speech), and 3) suspicions that the participant was not wearing the EAR. After these files were removed, participants were left with, on average, 209.5 valid audio files (69.8% of total; $SD = 57.0$, median = 220.0, range = 44-318 files). Valid files represented cases in which participants were awake and wearing the EAR, regardless of whether any speech was captured. Further, participants' speech was captured in 75.9 audio files on average ($SD = 32.0$, median = 73, range = 15-169 files with speech). We calculated the proportion of a participant's files with speech (see [Figure 2](#)), with participants speaking in anywhere between 8.2% to 76.1% of their valid files ($M = 37.1%$, $SD = 13.4%$, median = 34.9%). These numbers are largely consistent with other EAR datasets with college students (Manson & Robbins, 2017; Mehl et al., 2006; Mehl & Holleran, 2007; Mehl & Pennebaker, 2003).

3.3. Contexts of Speech

From the EAR, we are able to capture information regarding the contexts in which participants spoke more or less frequently. This data can shed light on the when, where, and with whom speech occurs. For example, do speakers generally communicate with individuals with high or low levels of common ground (shared knowledge/experience)? And what sorts of events (e.g., communicating with strangers in commercial settings) dominate day-to-day speech? Individuals can design their utterances for a specific listener (e.g., Clark & Murphy, 1982) but laboratory settings suggest that speakers are better at this utterance tailoring in some contexts than others (Gann & Barr, 2014). Understanding when speakers interact with individuals with whom they share knowledge, such as a known individual, versus someone with whom they have little shared knowledge, such as a stranger, may shed light on the kinds of experiences that speakers have accommodating listener knowledge, which is important for the interpretation of lab-

based behavior in which speakers succeed or fail at taking their listener's knowledge into account when designing utterances.

Many details about an individual's location are evident from audio cues alone. [Table 2](#) illustrates the locations in which participant speech was captured across all files that contained participant audio. The "home" category included the participant's dormitory, apartment, or other residence, or the home of a friend or family member. These percentages are roughly consistent with other EAR datasets collected from undergraduate populations and thus serve as a replication of these existing counts (Mehl & Pennebaker, 2003). As all participants were undergraduates, it may seem counterintuitive that speech in the classroom only less than 5% of total participant speech, and speech in the home accounted for nearly 60%. However, half of the recording period fell over the weekend, and in many cases, participants rarely left their homes during that time. In addition, many college classes are lecture-based, which does not allow for much speech production. Similarly, information about who the participant is talking to can be determined via the content of the speech being recorded, acoustic properties of the speech, and other auditory cues captured by the EAR. [Table 2](#) also details who the participants were talking to across all files containing participant speech. Participants overwhelmingly spoke with known individuals, and only rarely spoke to strangers. While some aspects of the EAR methodology (e.g., not wearing the device at work) may underestimate the prevalence of some conversations with strangers, overall, it may be the case that college-aged adults rarely converse with individuals with little shared knowledge or experiences, which has implications for the interpretation of young adult social behaviors in the lab.

3.4. Total Speech in English and Other Languages

After transcribing the speech of all audio files, we counted the total number of words uttered by each participant. Words were counted as they appear in the text, with the exception that English contractions were split at the apostrophe to yield two separate words. Little is known

Table 2. The percentage of audio files with participant speech recorded in various locations (left). The percentage of files in which the participant spoke to various individuals (right). Standard deviations are in parentheses. Percentages do not sum to 100% because some files may fall under more than one category.

Location	Speech Percentage	Interlocutor	Speech Percentage
Apartment/Dorm	55.6% (21.5)	To Self	7.2% (11.9)
Outdoor	6.9% (6.3)	To Known Person	87.4% (14.0)
Classroom	4.7% (8.0)	To Stranger	3.5% (5.1)
In Transit (Vehicle)	8.9% (8.6)	To Child	2.4% (6.1)
In Transit (Other)	6.4% (8.1)	To Pet	0.9% (2.3)
Bar/Coffee Shop/Restaurant	5.5% (6.0)	No Information	2.2% (4.5)
Shopping	2.2% (3.2)		
Other Public Place	11.2% (14.1)		
No Information	3.7% (9.7)		

about the day-to-day language habits of young adults, so this naturalistic data informs how individuals use spoken language in their daily lives. Further, there are many ways to quantify absolute or relative amounts of speech produced by different individuals, so these analyses explore and compare multiple approaches.

Of the 75 participants, 73 reported knowledge of more than one language, and 57 produced speech in a language other than English. Participants used their non-English language in 0-97.8% of their audio files ($M = 13.6\%$, median = 4.4%), and code-switched (used more than one language in a single file) in 0-31.0% of their files ($M = 6.4\%$, median = 2.9%). There is enormous variability across individuals in the frequency and contexts in which bilingual/multilingual speakers use English and other languages, and the EAR may be one means toward capturing some of that variability.

Computing word counts for non-English languages in a way that would allow a direct comparison to English presented some challenges. First, most code used to count words can only be used in languages that are written with spaces between words. Second, deciding what constitutes a single word versus multiple words across languages is not a trivial decision. In fact, some accounts question the psychological reality of a “word” representation more broadly (e.g., Baayen et al., 2016). Third, there is variability across languages in whether a given concept is realized as a single word or multiple words (e.g., “firetruck” in English versus “camión de bomberos” in Spanish). Fourth, there is variability across languages with respect to whether speakers can omit arguments like grammatical subjects or objects from utterances. Ideas, phrases, or sentences are frequently conveyed with different numbers of words in different languages. In an attempt to overcome the challenges associated with comparing different languages, we chose to compare English and non-English utterances in two different ways. First, we computed word counts for the non-English languages by calculating the number of words in the English translation of the utterance. While this method is not ideal because it is admittedly English-centric, it partially solved some of the concerns listed above. This method also captured variance in length across participant utterances, in a set of languages that varied across multiple typological and orthographic dimensions. Second, we counted the number of files that contained speech in other languages. While this

method ignores the length of the utterances captured in each recording, it avoids issues related to translation quality, language morphology, or other language features.

The two methods for assessing non-English use provide converging estimates (Figure 3). There was a strong correlation between the proportion of non-English speech files and the proportion of total non-English words (left panel; $r(55) = 0.96$, $p < 0.001$) and between the absolute numbers of files and words (right panel; $r(55) = 0.92$, $p < 0.001$), though removing a single outlier with a large amount of non-English speech yielded a somewhat weaker relationship ($r(54) = 0.79$, $p < 0.001$). Simply counting the number or proportion of files with non-English speech may be a valid means of calculating amounts or proportions of non-English speech. However, the distribution of non-English speech was highly skewed in our population, and the relationship between non-English audio files and words may be less reliable for samples with limited variability.

Figure 4 visualizes participants’ English and non-English use. The top panel shows the number of words produced in both English and another language, and the bottom panel shows the number of audio files that contained English or another language. Files that contained two languages are included in the counts for both languages. The most obvious feature of Figure 4 is variability in the amount and proportions of English and non-English speech. While some of the variability in overall amount of speech could be due to issues relating to EAR compliance, the proportion of valid files was only weakly (though significantly) related to the total number of words uttered across participants ($r(73) = 0.33$, $p < 0.01$). The number of valid files was more strongly related to the number of files containing speech ($r(73) = 0.49$, $p < 0.001$). Though a relation between EAR compliance and amount of speech captured is not ideal from the perspective of interpreting individual differences in language use, it is not surprising. However, the strength of the relation ($R^2 = 0.11$ and 0.24) suggests that many other factors other than EAR compliance contribute to the amount of speech produced.

From the number of words captured by the EAR, it is possible to estimate the total number of words spoken each day. By extrapolating from the number of total words (across all languages) and the number of valid files obtained from each participant and accounting for eight hours of

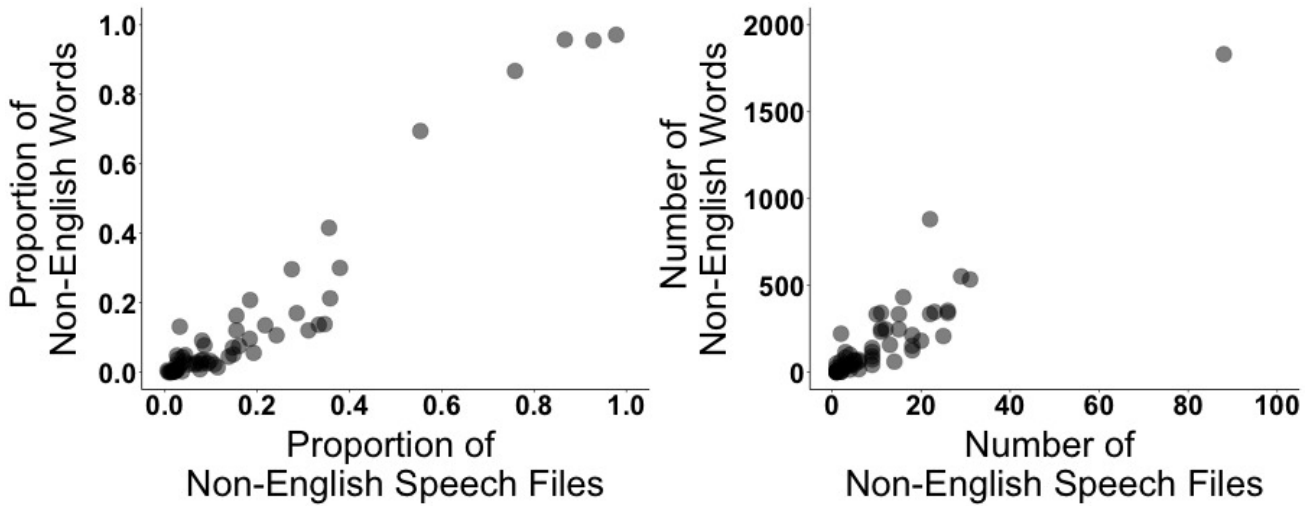


Figure 3. Relationships between the proportion of files from 57 participants containing non-English speech and the proportion of non-English words produced (left) and the total number of files containing non-English speech and the number of non-English words produced (right).

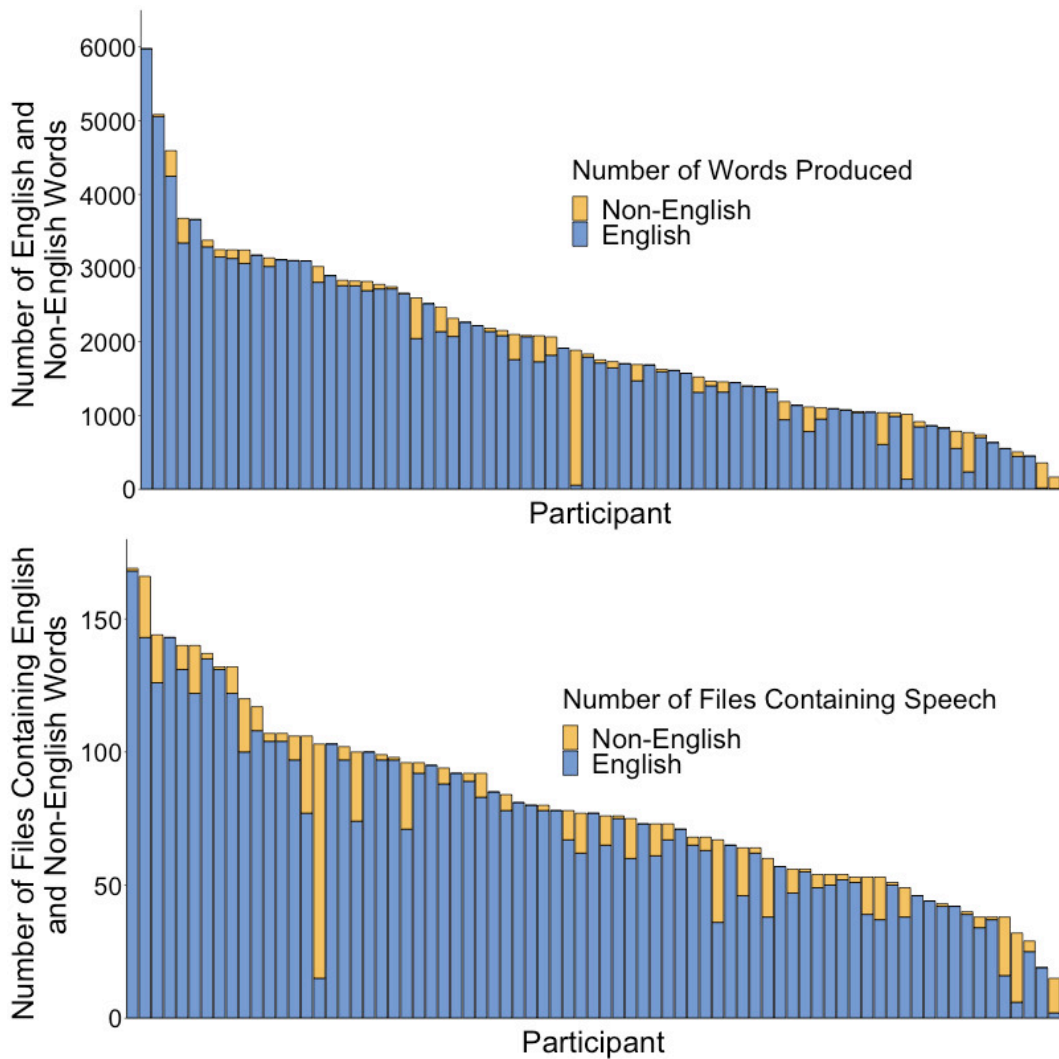


Figure 4. Count of all words uttered (top) and files containing speech (bottom) of the 75 participants.

sleep per night, we estimate that the average number of words produced by participants in our sample is approximately 14,096 words per day ($SD = 7,322$ words; range = 1,298 - 32,971 words). This figure is consistent with other EAR studies, which estimate that university students produce about 16,000 words per day (Mehl et al., 2007). For reference, 15,000 words is approximately 50 double-spaced APA-style pages of text and about half the length of Shakespeare's *Hamlet*. These results suggest that the number of words spoken per day by American college students may be highly variable and a potentially important source of variability in language experience.

3.5. Transcribing Audio versus Counting Files with Speech

Next, we investigated the necessity of fully transcribing audio files. For example, if the goal of the EAR data collection is simply to provide relative estimates of productive language (much like how estimates of spoken language input are often used in the language development literature), or assess relative use of two languages, is full transcription necessary? We investigated whether the total number of audio files that contain speech (a variable that is far less time-consuming to code) may be an appropriate proxy for the total amount of produced speech.

Overall, the total number of words (including all languages) in the fully transcribed dataset was correlated with the total number of audio files that contained participant speech ($r(73) = 0.85$, $p < 0.001$; Figure 5). A similar correlation was present when considering only English speech and English audio files ($r(73) = 0.87$, $p < 0.001$). Of course, when determining whether the *total number of transcribed words* and the *total number of audio files containing speech* assess the same construct, the relevant statistic is not whether these scores are significantly correlated, but the magnitude of the relation. Individual research questions and theoretical considerations will dictate whether a correlation coefficient of 0.85 ($R^2 = 0.72$) is an appropriately large magnitude to consider these measures to be similar. Research questions and theoretical motivation will also inform whether the total number of words or audio files with speech might be the more relevant variable to use. For example, utterance length may be less relevant if assessing time spent in environments with spoken language, rather than an individual's own speaking habits. Whether one measure or another better predicts a particular outcome variable is both an empirical and theoretical question.

To better understand the reasons underlying the divergence between the number of files containing speech and the total number of words produced, we also investigated the number of words produced in each audio file containing speech. As shown in the left panel of Figure 6, the total number of words produced (English and other languages collapsed together) was correlated with the average number of words contained in each audio file with speech ($r(73) = 0.67$, $p < 0.001$). A similar correlation was present when considering only English speech and audio files ($r(73) = 0.73$, $p < 0.001$). Both the number of files with speech and number of words per file contribute to the total number of words produced. Put simply, there seems to be multiple routes to pro-

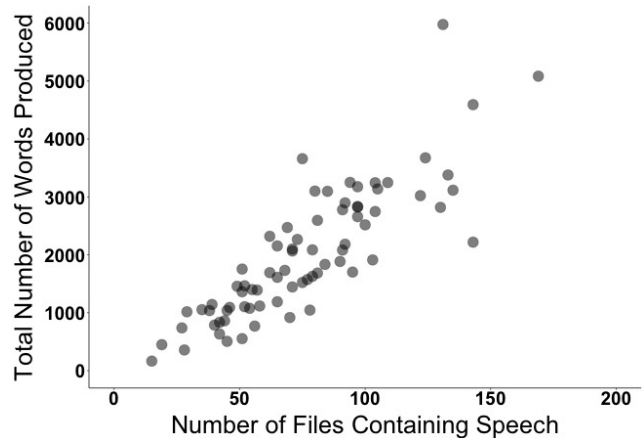


Figure 5. For all 75 participants, the relationship between the total number words uttered in all transcribed audio and the total number of audio files that contained participant speech.

ducing large amounts of speech. Individuals who produced more speech overall spoke more often and also tended to say more when they spoke. However, as shown in the right panel of Figure 6, the number of files with speech and the average number of words per file were only weakly correlated ($r(73) = 0.24$, $p < 0.05$), suggesting that it may not be the same individuals who spoke more often and produced more words when they spoke. This dissociation between speaking often and producing many words in a single utterance provides some explanation for the divergence between number of files containing speech and the total number of words produced.

3.6. Weekday versus Weekend Speech

Recording two weekdays and two weekend days of audio allowed us to compare language use when participants may encounter different speakers or speak in different contexts. Part of documenting individuals' language habits is documenting differences between the weekday and the weekend, which may contain vastly different profiles of speech. In particular, we hypothesized that our bilingual participants may interact with different individuals on weekdays and weekends (e.g., visiting family on weekends) and thus be more likely to speak their non-English language on weekends.

First, we saw no global tendency for participants to be more likely to wear the EAR during the week or weekend. We observed no difference overall between the number of valid files recorded ($t(74) = 0.68$, $p = 0.50$) or in the total number of files containing speech ($t(74) = 0.68$, $p = 0.50$). Figure 7 shows the distribution across all participants of the difference in the number of valid files (an approximate measure of EAR compliance) and the number of files with speech on weekdays and weekend. Individuals varied in their tendencies to wear the EAR and produce language during the week or weekend, suggesting that capturing speech during both weekdays and weekends may be a useful approach to gathering sufficient data in a large sample.

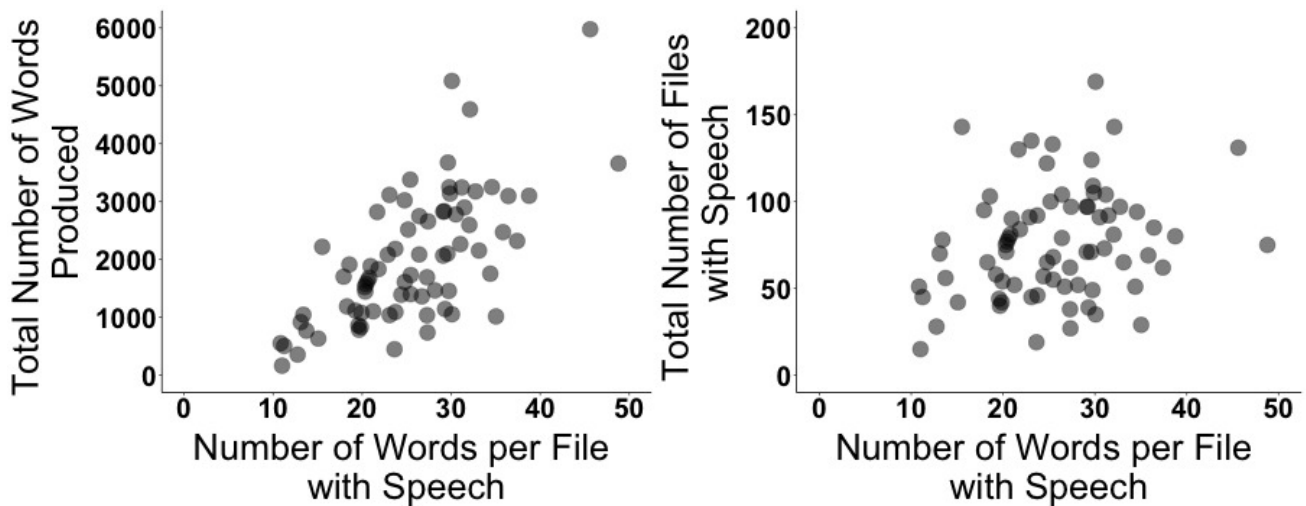


Figure 6. For all 75 participants, the relationship between the average number words uttered in each audio file containing speech and total number of words produced (left) and total number of audio files containing speech (right).

We predicted that bilingual students might use their two languages differently during the week and weekend, and this hypothesis was supported. Figure 8 illustrates the proportion of files with speech that included non-English speech. Participants are aligned in the top and bottom figure panel and ranked by overall proportion of non-English files. The blank column in the bottom panel refers to a single participant with no weekend speech files due to recording error. Among the 57 individuals who produced non-English speech, 15.8% of weekday files and 21.4% of weekend files contained non-English speech ($t(55) = 2.16, p < 0.05$). We found a greater proportion of non-English-containing speech files on the weekends. There may be considerable differences in patterns of speech during the week and weekend, so work investigating bilingual language use, especially in undergraduate populations, may want to consider possible differences in language use on different days of the week in order to assess an accurate snapshot of an individual's language habits.

3.7. Individual Differences in Lexical Diversity of Speech

In addition to the total amount of speech produced by participants, the EAR allowed us to examine the number of unique words produced by participants as a simple measure of lexical variability in speech. Figure 9 shows the number of total word tokens and unique word types produced by speakers, with separate counts for English and other languages that were spoken. While there is some individual variability in the lexical diversity of the speech (the vertical spread in points at a given sample size), what is most evident is the consistent relationship between total words and number of unique words across all participants. This is a relationship that is characteristic of coherent language (e.g., Malvern et al., 2004; Montag et al., 2018; Richards, 1987) and suggests that the total number of unique words or type-token ratios so strongly depend on sample size such that they are not appropriate measures of lexical diversity in

samples that vary in size. It may be unintuitive that there is little variability in the individuals' lexical diversity. Basic features of natural language—high frequency function words (e.g., *the, to, it*) must consistently appear alongside content words, conversations must be topically coherent, and many other cognitive or pragmatic constraints on spontaneous speech—limit the potential lexical diversity of daily speech. Lexical diversity may not be a reliable individual difference and may be driven by necessary features of natural language, rather than by vocabulary size or other individual differences.

3.8. Aggregate Lexical Statistics

In addition to quantifying aspects of individual participants' language use, the EAR can also be used to build a corpus of college students' spoken language by aggregating across many participants. This corpus could potentially be used to quantify normative aspects of college students' spoken language experience that could be used to predict various language behaviors. As mentioned previously, corpora that are used to describe and predict adult behavioral data often derive from written sources, which may present an incomplete or unrepresentative sample of an adult's language experience. The EAR may be a way to collect naturalistic samples of spoken language that can enable researchers to construct corpora that include spoken language experience.

Table 3 provides a small sample of the words contained in all participants' English audio recordings. The Subtlex data (51.0 million words) is from Brysbaert and New (2009). The Wikipedia data (approximately 2 billion words) was downloaded from Wikipedia (https://en.wikipedia.org/wiki/Wikipedia:Database_download). The COCA data (950 million words) was retrieved from the Corpus of Contemporary American English website (Davies, 2008-). The reddit data consists of 1.44 million words¹ from subreddits frequented by California college students. These corpora are all different sizes and consist of very different samples of language, produced for very different purposes and audiences. Given

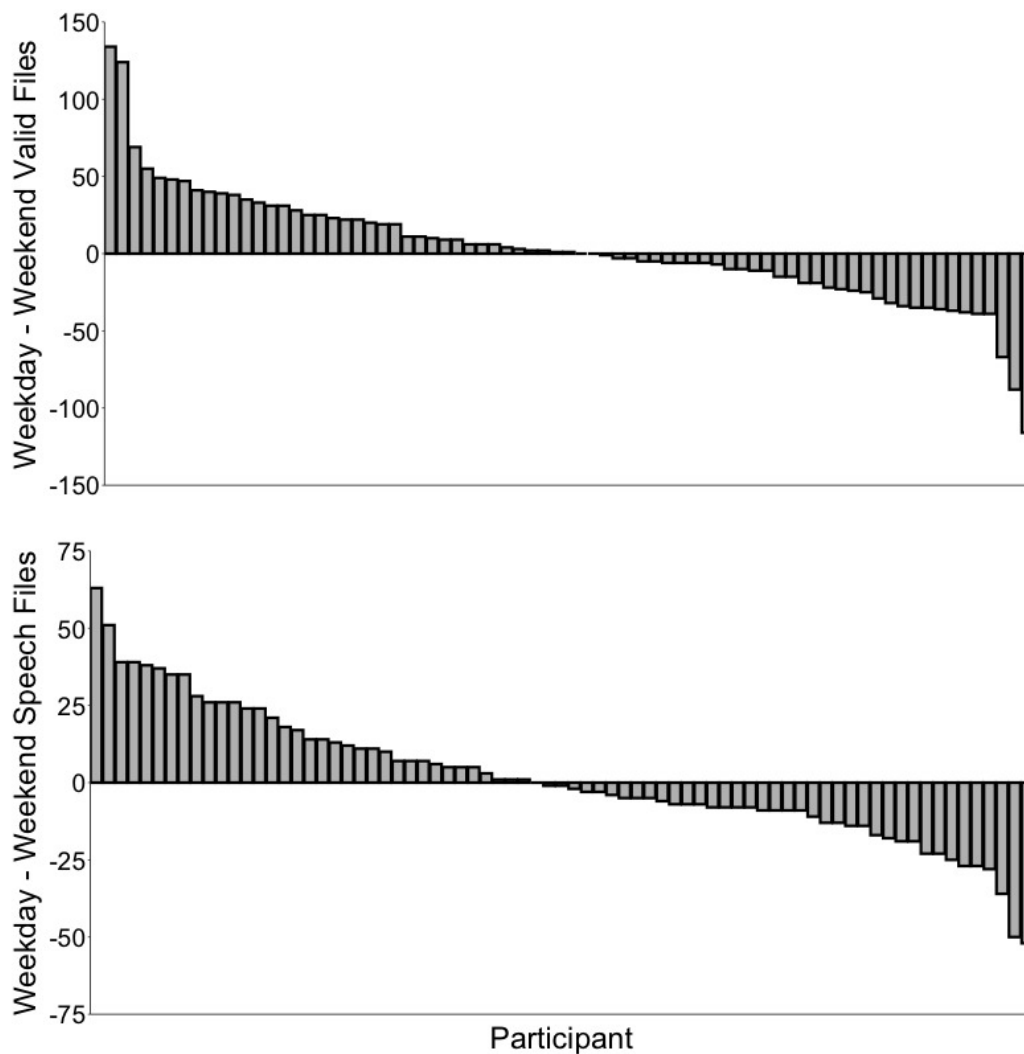


Figure 7. Difference in valid audio files (top) and audio files containing speech (bottom) on weekdays and weekends in all 75 participants. Positive values refer to a greater number of weekday files.

enormous differences in corpus size, measures like lexical diversity, or word proportions (e.g., the proportion of the corpus comprised of pronouns) are somewhat more complicated to compare across corpora (e.g., Malvern et al., 2004; Montag et al., 2018; Richards, 1987), but a list of the most frequent words can be compared across multiple corpora.

While there are obvious similarities across all five corpora with respect to the sets of words in each list, there are also subtle differences. For example, in the Wikipedia, COCA, and reddit corpora, which are dominated by written texts (COCA contains some scripted or semi-scripted spoken language), the most frequent word is “the.” By contrast, with EAR—which is spoken—and Subtlex—which is a movie subtitle corpus that is written but intended to approximate the spoken language register—“you” and “I” are among the

most frequent words. Of course, notable differences exist between the spontaneous spoken EAR corpus and Subtlex; fillers and discourse markers such as “like,” “yeah,” and “oh” are absent in Subtlex. The key finding from the comparison of words contained in different language samples is that the statistics in the EAR corpus may represent a different slice of language experience that can improve the representativeness of language corpora.

A list of all English words (133,159 total words) that appeared in the EAR corpus of all participants’ speech and their frequencies, along with two measures of contextual diversity (number of speakers who produced each word and number of different words that appear in a 7-word window around each word) is included in an online supplement (<https://osf.io/mpn4x/>).

1 Retrieved 9/28/2020. The Python package praw was used to scrape text from the 1,000 most recent posts and comments on the following subreddits: r/berkeley, r/ucla, r/ucr, r/UCL, r/UCSC, r/USCD, and r/UCSantaBarbara, or the most recent 6 months’ worth of posts, whichever came first. Only text posts and comments were used; posts of images and videos and their associated comment chains were not included.

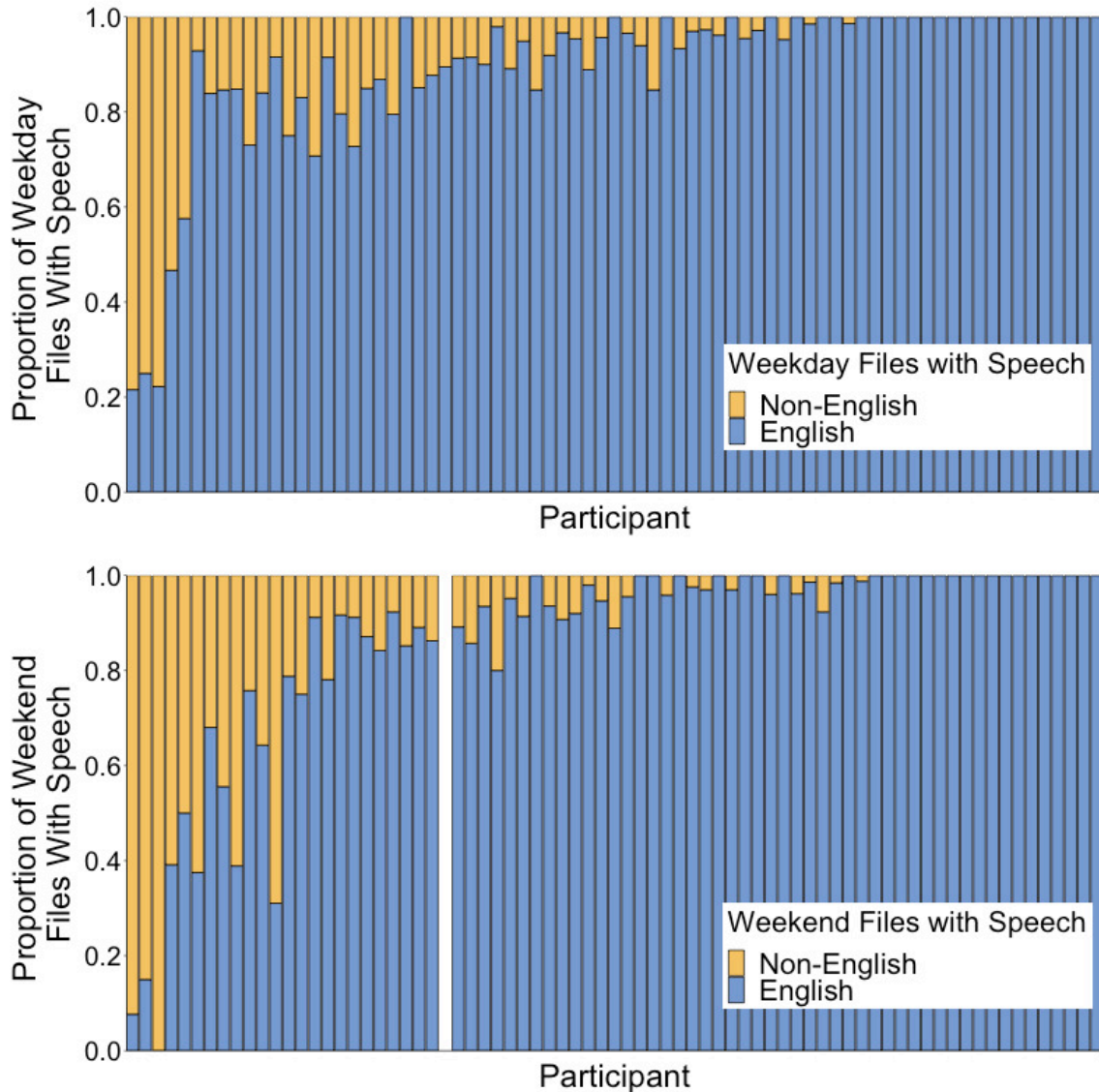


Figure 8. Proportion of weekday (top) and weekend (bottom) speech files containing non-English speech. All 75 participants are included, though only 57 participants' files contained non-English speech.

4. Discussion

We used the EAR as a means of collecting data about the day-to-day productive language experiences of American college students. We found that participants overwhelmingly spoke to individuals they knew, and produced about 14,000 words per day, though there was considerable variability across individuals. We also found that there were multiple routes to being a prolific speaker: speaking often, or speaking in long utterances, with the same individuals not necessarily exhibiting both tendencies. We also gained insight into the language habits of bilingual individuals. Many but not all of our participants spoke English more than they spoke another language, but most speakers who did speak in more than one language, used their non-English language more often on the weekend. When analyzing the content of participant utterances, we found little variability in the lexical diversity of utterances, likely owing to cognitive and pragmatic constraints on natural language.

We also present word frequency and lexical diversity counts of the words in our spoken corpus, which we expect will vary from, and complement, existing counts that derive from largely written corpora.

Time sampling, which involves collecting data or making observations at random or fixed intervals, is a powerful methodology frequently employed in the psychological sciences. The EAR is one example of time sampling that allows researchers to collect short audio samples across longer time frames to construct a picture of occurrences over that interval. We used the EAR as a means to collect information about adults' language habits that might be used to predict self-report measures of language use and behavior in other language tasks. We believe we have expanded the potential of the EAR to a new domain as a method of using time sampling to better understand the diverse language experiences of college students and other adults. Despite language experience, broadly, being an important aspect of studies of language use, little is known about patterns of adults' spo-

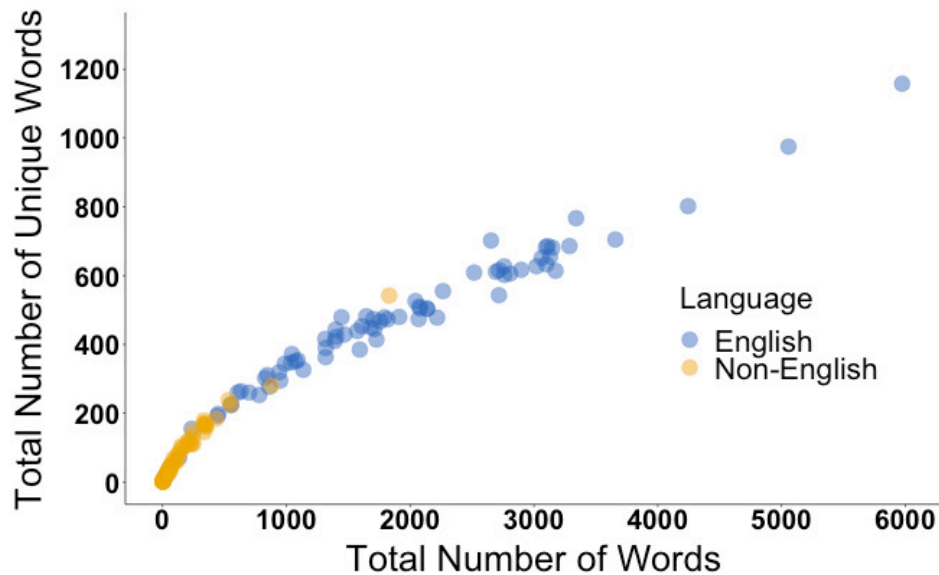


Figure 9. Number of total word tokens and unique word types produced by the participant spoke multiple languages (57 participants), those languages are indicated with separate points for English and their non-English language.

Table 3. A list of the 20 most frequent words in 5 different corpora.

Rank	EAR	SUBTLEX _{US}	WIKIPEDIA	COCA	Reddit
1	I	you	the	the	the
2	you	I	of	be [#]	I
3	it	the	in	and	to
4	like*	to	and	a	and
5	's	's	a	of	a
6	the	a	to	to	you
7	that	it	was	in	it
8	yeah	't	is	I	of
9	to	that	for	you	in
10	and	and	on	it	for
11	't	of	as	have [#]	is
12	oh	what	with	to	that
13	a	in	by	that	if
14	what	me	he	for	't
15	was	is	's	do [#]	but
16	so	we	that	he	on
17	my	this	at	with	be
18	is	he	from	on	have
19	no	on	his	this	my
20	know	for	it	n't	's

* Refers specifically to use of "like" as a filler or disfluency, not as a verb or comparison.

COCA is lemmatized, so these verbs refer to all conjugated forms. For example, *be* includes *is, are, was, 's* (when used as a contraction), and other forms of the verb *be*.

ken language and the types of language production and comprehension experiences that adults encounter through spoken language.

The present work highlights a number of advantages of using the EAR to understand adults' spoken language

habits. First, compliance among our participants was reasonably high, and in line with past research done in young adult populations. A noncompliance rate of only 8.4%, given the nature of what participants are being asked to do (wear an audio recorder and allow it to capture snippets of

your daily life at random intervals), and especially among undergraduates, is quite good. Further, a 79–81% average “wear rate” among the participants who did comply suggests that the request to wear the device over the course of four days was not too burdensome.

Turning to the content of the EAR recordings, a hallmark feature of our findings is the enormous amount of individual variability in different aspects of participant speech. This variability was evident in the total amount of speech produced, as measured by the number of audio files with speech or the number of words produced, the number of unique words spoken, and the frequency of weekday versus weekend speech. We find even more variability among our speakers who spoke more than one language in the total amount and proportion of their non-English language that they spoke, and when it was spoken. This variability in productive language suggests that summary measures stemming from corpora may not capture the nuances of individual speech patterns. Individual difference measures could also be used to capture variability in lexical complexity, syntactic complexity, rates of speech errors or disfluencies, or other features of participant speech. These measures of individual variability may be an important experience-based individual difference variable for predicting behavior on a range of lab-based language tasks, especially when the goal is to understand the role of language production experience on these tasks. Likewise, these utterances become language input to other speakers, so understanding patterns of speech in EAR data may serve not only as a measure of what adults *say* but also as a measure of what adults *hear*.

Related to this notion of variability is the observation that there were multiple routes to producing many words. Individuals who spoke often were not always the individuals who were saying a lot. In other words, an individual may have many audio files with speech, but few words in each file, whereas another individual may have few audio files with speech, but they say quite a bit within the 40-second intervals. This dissociation of audio files with speech and utterance length suggests another pattern in the way people speak that has not been subject to much investigation in the language literature. Likewise, it is unknown whether these linguistic profiles predict other aspects of behavior, perhaps in the social or personality domains.

Another source of variability encountered in the data was the contexts in which participants were speaking, as well as whom they were speaking to. Interestingly, much of the participants’ speech was captured in the home or a home-like setting, and participants were most often speaking to people that they knew. While our sample was limited in the sense that all participants were undergraduate students who may have had limited opportunities to leave campus, if this finding generalizes to other adult populations, much of a person’s everyday language may take place in the home or among people they know well. The extent to which speech is directed toward known individuals, with whom the speaker shares knowledge and common ground versus strangers with whom the speaker may not share much knowledge may be a factor that contributes to individual differences in the ability to successfully take an interlocutor’s perspective in lab-based tasks and may be worth investigating further.

In addition to highlighting individual differences in spoken language experience, other applications of the audio captured by the EAR may be to develop corpora of naturalistic speech. Most corpora used to describe and predict adults’ language behavior are written and may not capture distributional features of the spoken language that adults encounter. We saw this exemplified in our own comparison of our EAR corpus to other well-known corpora; first- and second-person pronouns (“I” and “you”) were much more common in the EAR corpus compared to written corpora, as well as a much greater occurrence of filler words such as “like.” Thus, it is clear that written language is not an adequate proxy for one’s overall language experience, and the EAR can be a tool by which researchers build corpora that *are* highly representative of adults’ spoken language experience. Further, there is a large amount of language experience that adults likely encounter via speech; we found that college students produced on average 14,000 words per day, but receptive spoken language is likely much more than this given that many speaking events have multiple listeners. As such, the inclusion of spoken language in corpora that predict adult language behavior may lead to better predictions. The time sampling aspect of the EAR is particularly well-suited for corpus construction because multiple conversational contexts can be sampled to produce more representative counts. We include word frequency counts and contextual diversity counts in our online materials and would be willing to add in measures from other research groups willing to share their data with the broader community.

Another unique aspect of the present work is that our sample consisted of many bilingual individuals who frequently spoke languages other than English. Outside of self-report surveys, little is known about the day-to-day language experiences of bilingual speakers, especially heritage bilingual speakers who mostly comprised our sample. The language experience of many individuals in the United States (and certainly around the world) may not be a monolingual one; thus, an individual’s lived language experience may not be adequately represented by existing accounts of spoken language experience. We demonstrated that the EAR can capture variability in the amount or type of experience that speakers have with multiple languages, to better quantify the ways in which a bilingual/multilingual speaker uses their languages. There is increased acknowledgement that there is enormous variability with respect to when, with whom, and how much bilingual speakers use their languages, and using the EAR to understand this variability may be crucial for adjudicating between different theories and hypotheses about bilingual language use (Gollan et al., 2015; Kroll et al., 2018). In future work, we will explore the relation of various individual variables to self-report measures of language use and measures of lab-based language behavior (Macbeth et al., 2021).

4.1. Future Directions of Transcription Technology

While the analyses we present here are fairly “low-tech” and represent a first-pass analysis of the rich information contained in the audio files, many new software or com-

putational modeling methodologies may become useful for analyzing audio data collected with the EAR. While current speech-to-text transcription tools are likely not yet sophisticated enough to handle the noisy audio recorded with EAR devices, automated or partially-automated transcription may be possible in the near future. However, in the meantime, it is possible that a classifier model could be built to distinguish audio files with speech in one language versus another language (e.g., English vs. Spanish speech) to rapidly count—with little or no human coding—the number of audio files containing speech of various languages. Given that for some research questions the number of files with speech may be an appropriate proxy for the number of words produced, classifier models such as these could be a way to dramatically speed data annotation time, with possible implications for participant privacy because no human would need to listen to the audio recordings.

Software, especially various Natural Language Processing software packages, could also be used to analyze the audio recordings or transcribed text from the EAR devices. For example, Linguistic Inquiry and Word Count (LIWC) has often been used to analyze the linguistic content of EAR transcripts (Mehl et al., 2001; Mehl & Pennebaker, 2003; Robbins et al., 2019), as have other Python-based software packages (Luo, Robbins, et al., 2019; Luo, Schneider, et al., 2019). With the increase in popularity and quality of other Natural Language Processing tools—including sentiment analyses, sentence parsers, part of speech taggers, and many others—additional analyses could be performed. However, some analyses may not be possible because of 1) sample sizes potentially being small and 2) full, topically-coherent conversations rarely being captured due to the time sampling nature of the EAR. Nonetheless, there are many future avenues to use new Natural Language Processing techniques on this naturalistic speech to perform analyses of the linguistic content of the captured audio.

Finally, groups interested in specific timing details of the speech captured have a number of options. First, software packages such as ELAN (<https://archive.mpi.nl/tla/elan>; Wittenburg et al., 2006) may be helpful for segmenting speech by speaker, as well as for transcribing and annotating speech in a way that timestamps utterance boundaries to allow for various analyses of utterance or conversation timing to be performed. These tools may be particularly useful for audio recordings with multiple speakers or for capturing temporal dynamics of turn taking during conversations. Other timing analyses could be performed with forced aligners, such as the Montreal Forced Aligner (McAuliffe et al., 2017), which can compute the word-by-word timing of an utterance from the raw audio file and text transcript. Forced aligners may obviate the need to hand-code speaking latencies and durations, which can enable researchers to quickly and easily perform timing analyses of naturalistic speech after it has been transcribed.

4.2. Limitations of the EAR Methodology

In the course of using the EAR as a means of collecting language samples from a university population, we encountered both methodological and human subjects-related challenges. Some limitations arose from methodological

details associated with our particular study. About two-thirds of our sample was female, and thus, it is possible that our data may be more representative of female speech patterns. However, men and women in our dataset did not differ in their proportion of valid files with speech or total words spoken, suggesting that in general, men and women produce similar amounts of speech, consistent with past research (Mehl et al., 2007). More work will need to be done to assess whether gender differences exist in finer-grained aspects of word use in this population.

We also noticed important limits to the audio quality of the EAR, specifically that the audibility of speech drops off significantly as distance between the background speaker and the EAR increases. One aspect of language experience that we had been interested in capturing was a measure of passive exposure to non-English languages. We were interested in how often our undergraduate population encountered different languages in various contexts and how this exposure might affect their own language use. We initially attempted to code for whether there was language in the background (that was not directed at the participant), and if so, which language(s). Hypothetically, we could then calculate a proportion of audio files that contained ambient speech in different languages. Unfortunately, the poor audio quality of speech far from the EAR made it difficult for coders to determine which languages were present in the background.

Another important point is that the majority of our sample consisted of bi- or multilingual college students from Southern California, a linguistically diverse region of the United States. It is unclear the extent to which the speech patterns captured here are broadly representative of bilingual young adults, or young adult college students outside of Southern California or the United States. It is also possible that day-to-day language use might differ as a function of the sample age. Language evolves with each new generation, and certain words or phrases that are prevalent in a college-aged sample might not be among those used by middle-aged or older adults. We urge caution in generalizing the results presented here to other populations.

4.3. Conclusions

Ultimately, we advocate for the use of the EAR in cognitive and language science. We suggest that it might be particularly useful for language researchers who wish to examine spoken language in adult populations, because analyses of actual adult-to-adult spoken language are virtually nonexistent in today's psycholinguistics, as either an individual difference measure or as a means of building normative corpora. This naturalistic data also paves the way for a better understanding of how well lab-based linguistic tasks capture patterns of day-to-day language use, and how multilingual speakers use their languages in different ways.

Acknowledgments

We would like to thank all of our research assistants: Arpine Agakhanyan, Bani Brara, Pei Chai, Amber Culwell,

Isha Gupta, Verna Halim, Tzu-Ning Vicky Hsu, Mariamme Ibrahim, Nahleh Koochak, Jasmine Jahandar, Yuumi Amy Kobayashi, Tram Le, Vanessa Ledesma, Jamie Lee, Belen Leon, Ziomara Machado, Eanna Mejia, Monica Mikhail, Melissa Ramos, Cindy Sarabia, Daniya Siddiqua, Stephanie Silva, Supanat Sritapan, and Sanna Tahir, for their invaluable help with data coding and transcription.

Funding

This work was partially supported by a James S. McDonnell Foundation Scholar Award to JLM, as well as a National Science Foundation Postdoctoral Research Fellowship under Grant No. SBE-1714925 and CSUF Junior Grant to NA.

Competing Interests

The authors have no competing interests to declare.

Data Accessibility Statement

All data, coding materials, and analysis scripts can be found on this paper's project page at <https://osf.io/mpn4x/>.

Submitted: June 04, 2021 PDT, Accepted: May 20, 2022 PDT



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

References

- Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, *17*(9), 814–823. <https://doi.org/10.1111/j.1467-9280.2006.01787.x>
- Allemand, M., & Mehl, M. R. (2017). Personality assessment in daily life: A roadmap for future personality development research. In J. Specht (Ed.), *Personality development across the lifespan* (pp. 437–454). Elsevier.
- Anderson, J. A. E., Mak, L., Chahi, A. K., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods*, *50*(1), 250–263. <https://doi.org/10.3758/s13428-017-0867-9>
- Arnold, J. E., Strangmann, I. M., Hwang, H., Zerkle, S., & Nappa, R. (2018). Linguistic experience affects pronoun interpretation. *Journal of Memory and Language*, *102*, 41–54. <https://doi.org/10.1016/j.jml.2018.05.002>
- Baayen, R. H., Shaoul, C., Willits, J., & Ramscar, M. (2016). Comprehension without segmentation: A proof of concept with naive discriminative learning. *Language, Cognition and Neuroscience*, *31*(1), 106–128. <https://doi.org/10.1080/23273798.2015.1065336>
- Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283–316. <https://doi.org/10.1037/0096-3445.133.2.283>
- Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., Herron, D., Ching Lu, C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., ... Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*(2), 344–380. <https://doi.org/10.3758/bf03196494>
- Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A. S., & Amatuni, A. (2019). What do north American babies hear? A large-scale cross-corpus analysis. *Developmental Science*, *22*, e12724. <https://doi.org/10.1111/desc.12724>
- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511621024>
- Bloom, L., Lightbown, P., Hood, L., Bowerman, M., Maratsos, M., & Maratsos, M. P. (1975). Structure and variation in child language. *Monographs of the Society for Research in Child Development*, *40*(Serial No. 160). <https://doi.org/10.2307/1165986>
- Bollich, K. L., Doris, J. M., Vazire, S., Raison, C. L., Jackson, J. J., & Mehl, M. R. (2016). Eavesdropping on character: Assessing everyday moral behaviors. *Journal of Research in Personality*, *61*, 15–21. <https://doi.org/10.1016/j.jrp.2015.12.003>
- Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind*, *3*, 52–67. https://doi.org/10.1162/opmi_a_00026
- Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*(2), B33–B44. [https://doi.org/10.1016/s0010-0277\(01\)00122-6](https://doi.org/10.1016/s0010-0277(01)00122-6)
- Brown, R. (1973). *A first language: The early stages*. Harvard University Press. <https://doi.org/10.4159/harvard.9780674732469>
- Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, *58*(5), 412–424. <https://doi.org/10.1027/1618-3169/a000123>
- Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. <https://doi.org/10.3758/brm.41.4.977>
- Burgess, C., & Livesay, K. (1998). The effect of corpus size in predicting reaction time in a basic word recognition task: Moving on from Kučera and Francis. *Behavior Research Methods, Instruments, & Computers*, *30*(2), 272–277. <https://doi.org/10.3758/bf03200655>
- Cartmill, E. A., Armstrong, B. F., III, Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, *110*(28), 11278–11283. <https://doi.org/10.1073/pnas.1309518110>
- Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a Tzeltal Mayan village. *Child Development*, *91*(5), 1819–1835. <https://doi.org/10.1111/cdev.13349>
- Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In J. F. Le Ny & W. Kintsch (Eds.), *Language and comprehension* (pp. 287–296). North-Holland. [https://doi.org/10.1016/s0166-4115\(09\)60059-5](https://doi.org/10.1016/s0166-4115(09)60059-5)
- Clark, H. H., & Tree, J. E. F. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73–111. [https://doi.org/10.1016/s0010-0277\(02\)00017-3](https://doi.org/10.1016/s0010-0277(02)00017-3)
- Cunningham, A. E., & Stanovich, K. E. (1998). What reading does for the mind. *American Educator*, *22*, 8–17.
- Davies, M. (2008). *The Corpus of Contemporary American English (COCA)*. <https://www.english-corpora.org/coca/>
- de Villiers, J. G., & de Villiers, P. A. (1973). A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of Psycholinguistic Research*, *2*(3), 267–278. <https://doi.org/10.1007/bf01067106>

- Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, *104*(1), 123–147. <http://doi.org/10.1037/0033-295x.104.1.123>
- Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 611–629. [https://doi.org/10.1016/s0022-5371\(81\)90202-4](https://doi.org/10.1016/s0022-5371(81)90202-4)
- Demiray, B., Luo, M., Tejada-Padron, A., & Mehl, M. R. (2020). Sounds of healthy aging: Assessing everyday social and cognitive activity from ecologically sampled ambient audio data. In *Personality and healthy aging in adulthood* (pp. 111–132). Springer.
- Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, *152*, 101–107. <https://doi.org/10.1016/j.cognition.2016.03.005>
- Fisher, A., Reilly, J. J., Kelly, L. A., Montgomery, C., Williamson, A., Paton, J. Y., & Grant, S. (2005). Fundamental movement skills and habitual physical activity in young children. *Medicine & Science in Sports & Exercise*, *37*(4), 684–688. <https://doi.org/10.1249/01.mss.0000159138.48107.7d>
- Ford, M., Baer, C. T., Xu, D., Yapanel, U., & Gray, S. (2008). *The LENATM language environment analysis system: Audio specifications of the DLP-0121*. Lena Foundation.
- Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-mounted eye tracking: A new method to describe infant looking. *Child Development*, *82*(6), 1738–1750. <https://doi.org/10.1111/j.1467-8624.2011.01670.x>
- Fromkin, V. A. (1973). Slips of the tongue. *Scientific American*, *229*(6), 110–117. <https://doi.org/10.1038/scientificamerican1273-110>
- Gann, T. M., & Barr, D. J. (2014). Speaking from experience: Audience design as expert performance. *Language, Cognition and Neuroscience*, *29*(6), 744–760. <https://doi.org/10.1080/01690965.2011.641388>
- Garnham, A., Shillcock, R. C., Brown, G. D. A., Mill, A. I. D., & Cutler, A. (1982). Slips of the tongue in the London-Lund corpus of spontaneous conversation. In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 251–263). Mouton.
- Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, *37*(1), 58–93. <https://doi.org/10.1006/jmla.1997.2512>
- Gennari, S. P., & MacDonald, M. C. (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition*, *111*(1), 1–23. <https://doi.org/10.1016/j.cognition.2008.12.006>
- Gilkerson, J., & Richards, J. A. (2008). *The LENA natural language study*. LENA Foundation.
- Gilkerson, J., Richards, J. A., Warren, S. F., Oller, D. K., Russo, R., & Vohr, B. (2018). Language experience in the second year of life and language outcomes in late childhood. *Pediatrics*, *142*(4), e20174276. <https://doi.org/10.1542/peds.2017-4276>
- Godfrey, J. J., & Holliman, E. (1993). *Switchboard-1 Release 2 LDC97S62*. Web Download. Linguistic Data Consortium.
- Goldin-Meadow, S., Levine, S. C., Hedges, L. V., Huttenlocher, J., Raudenbush, S. W., & Small, S. L. (2014). New evidence about language and cognitive development based on a longitudinal study: Hypotheses for intervention. *American Psychologist*, *69*(6), 588–599. <https://doi.org/10.1037/a0036886>
- Gollan, T. H., Starr, J., & Ferreira, V. S. (2015). More than use it or lose it: The number-of-speakers effect on heritage language proficiency. *Psychonomic Bulletin & Review*, *22*(1), 147–155. <https://doi.org/10.3758/s13423-014-0649-7>
- Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, *35*(3), 515–531. <https://doi.org/10.1017/s0305000907008641>
- Hare, M., Tanenhaus, M. K., & McRae, K. (2007). Understanding and producing the reduced relative construction: Evidence from ratings, editing, and corpora. *Journal of Memory & Language*, *56*(3), 410–435. <https://doi.org/10.1016/j.jml.2006.08.007>
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experiences of young American children*. Brookes.
- Hayes, D. P. (1988). Speaking and writing: Distinct patterns of word choice. *Journal of Memory and Language*, *27*(5), 572–585. [https://doi.org/10.1016/0749-596x\(88\)90027-7](https://doi.org/10.1016/0749-596x(88)90027-7)
- Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, *63*(3), 259–273. <https://doi.org/10.1016/j.jml.2010.06.002>
- Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? *Psychological Science*, *20*(6), 729–739. <https://doi.org/10.1111/j.1467-9280.2009.02365.x>
- Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., Yust, P. K. S., & Suma, K. (2015). The contribution of early communication quality to low-income children's language success. *Psychological Science*, *26*(7), 1071–1083. <https://doi.org/10.1177/0956797615581493>
- Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*, *73*(2), 418–433. <https://doi.org/10.1111/1467-8624.00415>
- Huebner, P. A., & Willits, J. A. (2018). Structured semantic knowledge can emerge automatically from predicting word sequences in child-directed speech. *Frontiers in Psychology*, *9*, 133. <https://doi.org/10.3389/fpsyg.2018.00133>
- Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, *11*(6), F31–F39. <https://doi.org/10.1111/j.1467-7687.2008.00768.x>

- Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & et al. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, 27(2), 236–248. <https://doi.org/10.1037/0012-1649.27.2.236>
- Johns, B. T., & Jamieson, R. K. (2018). A large-scale analysis of variance in written language. *Cognitive Science*, 42(4), 1360–1374. <https://doi.org/10.1111/cogs.12583>
- Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, 114(1), 1–37. <https://doi.org/10.1037/0033-295x.114.1.1>
- Kaplan, D. M., Rentscher, K. E., Lim, M., Reyes, R., Keating, D., Romero, J., Shah, A., Smith, A. D., York, K. A., Milek, A., Tackman, A. M., & Mehl, M. R. (2020). Best practices for Electronically Activated Recorder (EAR) research: A practical guide to coding and processing EAR data. *Behavior Research Methods*, 52(4), 1538–1551. <https://doi.org/10.3758/s13428-019-01333-y>
- Karan, A., Wright, R. C., & Robbins, M. L. (2017). Everyday emotion word and personal pronoun use reflects dyadic adjustment among couples coping with breast cancer. *Personal Relationships*, 24(1), 36–48. <https://doi.org/10.1111/pere.12165>
- Kroll, J. F., Dussias, P. E., & Bajo, M. T. (2018). Language use across international contexts: Shaping the minds of L2 speakers. *Annual Review of Applied Linguistics*, 38, 60–79. <https://doi.org/10.1017/s0267190518000119>
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177. <http://doi.org/10.1016/j.cognition.2007.05.006>
- Li, P., Zhang, F., Yu, A., & Zhao, X. (2020). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, 23(5), 938–944. <https://doi.org/10.1017/s1366728918001153>
- Lieven, E. (2016). Usage-based approaches to language development: Where do we go from here? *Language and Cognition*, 8(3), 346–368. <https://doi.org/10.1017/langcog.2016.16>
- Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, 28(2), 203–208. <https://doi.org/10.3758/bf03204766>
- Luo, M., Robbins, M. L., Martin, M., & Demiray, B. (2019). Real-life language use across different interlocutors: A naturalistic observation study of adults varying in age. *Frontiers in Psychology*, 10, 1412. <https://doi.org/10.3389/fpsyg.2019.01412>
- Luo, M., Schneider, G., Martin, M., & Demiray, B. (2019). Cognitive aging effects on language use in real-life contexts: A naturalistic observation study. *41st Annual Conference of the Cognitive Science Society*, 714–720.
- Macbeth, A., Bruni, M., De La Cruz, B., Montag, J. L., Atagi, N., Robbins, M. L., & Chiarello, C. (2021). *How Real-World Language Use Relates to Self-Report and Laboratory Measures of Bilingualism* [Manuscript submitted for publication].
- MacKay, D. G. (1972). The structure of words and syllables: Evidence from errors in speech. *Cognitive Psychology*, 3(2), 210–227. [https://doi.org/10.1016/001-0285\(72\)90004-7](https://doi.org/10.1016/001-0285(72)90004-7)
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Lawrence Erlbaum Associates.
- Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Palgrave Macmillan UK. <https://doi.org/10.1057/9780230511804>
- Manson, J. H., & Robbins, M. L. (2017). New evaluation of the Electronically Activated Recorder (EAR): Obtrusiveness, compliance, and participant self-selection effects. *Frontiers in Psychology*, 8, 658. <http://doi.org/10.3389/fpsyg.2017.00658>
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, 50(4), 940–967. [http://doi.org/10.1044/1092-4388\(2007\)067](http://doi.org/10.1044/1092-4388(2007)067)
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. *Proceedings of the 18th Conference of the International Speech Communication Association*, 498–502.
- Mehl, M. R. (2017). The electronically activated recorder (EAR): A method for the naturalistic observation of daily social behavior. *Current Directions in Psychological Science*, 26(2), 184–190. <https://doi.org/10.1177/0963721416680611>
- Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, 90(5), 862–877. <https://doi.org/10.1037/0022-3514.90.5.862>
- Mehl, M. R., & Holleran, S. E. (2007). An empirical analysis of the obtrusiveness of and participants' compliance with the electronically activated recorder (EAR). *European Journal of Psychological Assessment*, 23(4), 248–257. <https://doi.org/10.1027/1015-5759.23.4.248>
- Mehl, M. R., & Pennebaker, J. W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology*, 84(4), 857–870. <https://doi.org/10.1037/0022-3514.84.4.857>
- Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., & Price, J. H. (2001). The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, & Computers*, 33(4), 517–523. <https://doi.org/10.3758/bf03195410>
- Mehl, M. R., Vazire, S., Ramírez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men? *Science*, 317(5834), 82. <http://doi.org/10.1126/science.1139940>
- Mendoza, J. K., & Fausey, C. M. (2021). Everyday music in infancy. *Developmental Science*, 24(6), e13122. <https://doi.org/10.1111/desc.13122>

- Minor, K. S., Davis, B. J., Marggraf, M. P., Luther, L., & Robbins, M. L. (2018). Words matter: Implementing the electronically activated recorder in schizotypy. *Personality Disorders: Theory, Research, and Treatment*, 9(2), 133–143. <https://doi.org/10.1037/per0000266>
- Montag, J. L. (2020). New insights from daylong audio transcripts of children's language environments. *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, 3005–3011.
- Montag, J. L., Jones, M. N., & Smith, L. B. (2018). Quantity and diversity: Simulations of early word learning environments. *Cognitive Science*, 42(S2), 375–412. <https://doi.org/10.1111/cogs.12592>
- Montag, J. L., & MacDonald, M. C. (2015). Text exposure predicts spoken production of complex sentences in 8- and 12-year-old children and adults. *Journal of Experimental Psychology: General*, 144(2), 447–468. <https://doi.org/10.1037/xge0000054>
- Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: New insights and technical developments. *World Psychiatry*, 17(2), 123–132. <https://doi.org/10.1002/wps.20513>
- Nisbett, R. E., & Miyamoto, Y. (2005). The influence of culture: Holistic versus analytic perception. *Trends in Cognitive Sciences*, 9(10), 467–473. <https://doi.org/10.1016/j.tics.2005.08.004>
- Oller, D. K., Griebel, U., Iyer, S. N., Jhang, Y., Warlaumont, A. S., Dale, R., & Call, J. (2019). Language origins viewed in spontaneous and interactive vocal rates of human and bonobo infants. *Frontiers in Psychology*, 10, 729. <https://doi.org/10.3389/fpsyg.2019.00729>
- Olney, A. M., Dale, R., & D'Mello, S. K. (2012). The world within Wikipedia: An ecology of mind. *Information*, 3(2), 229–255. <https://doi.org/10.3390/info3020229>
- Payne, B. R., Gao, X., Noh, S. R., Anderson, C. J., & Stine-Morrow, E. A. L. (2012). The effects of print exposure on sentence processing and memory in older adults: Evidence for efficiency and reserve. *Aging, Neuropsychology, and Cognition*, 19(1–2), 122–149. <https://doi.org/10.1080/13825585.2011.628376>
- Pereira, F., Gershman, S., Ritter, S., & Botvinick, M. (2016). A comparative evaluation of off-the-shelf distributed semantic representations for modelling behavioural data. *Cognitive Neuropsychology*, 33(3–4), 175–190. <https://doi.org/10.1080/02643294.2016.1176907>
- Pretzer, G. M., Lopez, L. D., Walle, E. A., & Warlaumont, A. S. (2019). Infant-adult vocal interaction dynamics depend on infant vocal type, child-directedness of adult speech, and timeframe. *Infant Behavior and Development*, 57, 101325. <https://doi.org/10.1016/j.infbeh.2019.04.007>
- Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2017). The impact of early social interaction on later language development in Spanish-English bilingual infants. *Child Development*, 88(4), 1216–1234. <https://doi.org/10.1111/cdev.12648>
- Real, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, 57(1), 1–23. <https://doi.org/10.1016/j.jml.2006.08.014>
- Recchia, G., & Jones, M. N. (2009). More data trumps smarter algorithms: Comparing pointwise mutual information with latent semantic analysis. *Behavior Research Methods*, 41(3), 647–656. <https://doi.org/10.3758/brm.41.3.647>
- Richards, B. (1987). Type/token ratios: What do they really tell us? *Journal of Child Language*, 14(2), 201–209. <https://doi.org/10.1017/s0305000900012885>
- Robbins, M. L. (2017). Practical suggestions for legal and ethical concerns with social environment sampling methods. *Social Psychological and Personality Science*, 8(5), 573–580. <https://doi.org/10.1177/1948550617699253>
- Robbins, M. L., Focella, E. S., Kastle, S., López, A. M., Weihs, K. L., & Mehl, M. R. (2011). Naturalistically observed swearing, emotional support, and depressive symptoms in women coping with illness. *Health Psychology*, 30(6), 789–792. <https://doi.org/10.1037/a0023431>
- Robbins, M. L., Karan, A., López, A. M., & Weihs, K. L. (2018). Naturalistically observing noncancer conversations among couples coping with breast cancer. *Psycho-Oncology*, 27(9), 2206–2213. <https://doi.org/10.1002/pon.4797>
- Robbins, M. L., López, A. M., Weihs, K. L., & Mehl, M. R. (2014). Cancer conversations in context: Naturalistic observation of couples coping with breast cancer. *Journal of Family Psychology*, 28(3), 380–390. <https://doi.org/10.1037/a0036458>
- Robbins, M. L., Wright, R. C., María López, A., & Weihs, K. (2019). Interpersonal positive reframing in the daily lives of couples coping with breast cancer. *Journal of Psychosocial Oncology*, 37(2), 160–177. <https://doi.org/10.1080/07347332.2018.1555198>
- Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, 57(3), 348–379. <https://doi.org/10.1016/j.jml.2007.03.002>
- Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language*, 35(1), 185–205. <https://doi.org/10.1017/s0305000907008343>
- Slone, L. K., Abney, D. H., Borjon, J. I., Chen, C., Franchak, J. M., Percy, D., Suarez-Rivera, C., Xu, T. L., Zhang, Y., Smith, L. B., & Yu, C. (2018). Gaze in action: Head-mounted eye tracking of children's dynamic visual attention during naturalistic behavior. *Journal of Visualized Experiments*, 141, e58496. <https://doi.org/10.3791/58496>
- Snow, C. E. (1977). The development of conversation between mothers and babies. *Journal of Child Language*, 4(1), 1–22. <https://doi.org/10.1017/s0305000900000453>
- Street, J. A., & Dabrowska, E. (2010). More individual differences in language attainment: How much do adult native speakers of English know about passives and quantifiers? *Lingua*, 120(8), 2080–2094. <https://doi.org/10.1016/j.lingua.2010.01.004>

- Svartvik, J., & Quirk, R. (Eds.). (1980). *A corpus of English conversation*. Gleerup.
- Swingle, D., & Humphrey, C. (2018). Quantitative linguistic predictors of infants' learning of specific English words. *Child Development, 89*(4), 1247–1267. <https://doi.org/10.1111/cdev.12731>
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*(1), 24–54. <http://doi.org/10.1177/0261927x09351676>
- Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(3), 528–553. <https://doi.org/10.1037/0278-7393.19.3.528>
- VanDam, M., Warlaumont, A. S., Bergelson, E., Cristia, A., Soderstrom, M., De Palma, P., & MacWhinney, B. (2016). HomeBank: An online repository of daylong child-centered audio recordings. *Seminars in Speech & Language, 37*(02), 128–142. <https://doi.org/10.1055/s-0036-1580745>
- Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., & Miller, J. (1985). From babbling to speech: A re-assessment of the continuity issue. *Language, 61*(2), 397–445. <https://doi.org/10.2307/414151>
- Wank, A. A., Mehl, M. R., Andrews-Hanna, J. R., Polsinelli, A. J., Moseley, S., Glisky, E. L., & Grilli, M. D. (2020). Eavesdropping on autobiographical memory: A naturalistic observation study of older adults' memory sharing in daily conversations. *Frontiers in Human Neuroscience, 14*, 238. <https://doi.org/10.3389/fnhum.2020.00238>
- Willits, J. A., Seidenberg, M. S., & Saffran, J. R. (2014). Distributional structure in language: Contributions to noun-verb difficulty differences in infant word recognition. *Cognition, 132*(3), 429–436. <https://doi.org/10.1016/j.cognition.2014.05.004>
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A Professional Framework for Multimodality Research. *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*.
- Zevin, J. D., & Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language, 47*(1), 1–29. <http://doi.org/10.1006/jmla.2001.2834>

Appendices

Appendix A: Intraclass Correlation Coefficients Across Coding Categories

We assessed coder reliability by calculating the intraclass correlation coefficient (ICC) for each of the codes using a one-way random effects model. All coding categories reference the participant, except for the last two which reference conversation partner language use. Only a few codes were removed due to low reliability when averaged across waves: whether the participant was doing housework (.545) or socializing (.267), the conversation partner's non-English language (.448), and the presence of language(s) in the background (.534 for first language heard, and .539 for second language heard, if applicable). It was difficult for coders to reliably identify these activities/features from the audio alone. For each of our remaining codes, the average ICC across both waves was above .60, and the average ICC for all codes across the two phases was .87. This corroborates ICC calculations from other EAR studies (Karan et al., 2017; Robbins et al., 2014, 2018).

Appendix B: Recommendations for Transcription and Computer Code

Many groups may choose to use Python or another programming language to automate various counting and analyses of the transcribed audio. We discovered a number of features that the audio transcribers may want to include—or not include—to make data analysis automation easier. First, the transcripts often require a number of codes to denote different features of the audio. For example, often it is necessary to include codes in the transcript to indicate which language is being spoken or features of the background audio. We suggest that these codes not be letters or strings of letters that appear in other words, so that these codes can easily be searched or removed without altering other words in the transcript. Researchers may also want to make these codes similar to each other to streamline the regular expressions that need to be used to automatically count, remove, or replace the codes. Second, if groups intend to use automatic tagging and parsing programs that strip punctuation-based markers from text, they may want to avoid using punctuation markers in their codes because those codes could be altered by the program. Alternately, avoiding automatic parsing (like in spacy) and writing the code oneself would be another way around this problem and allow for a wider range of punctuation to be used as codes. Third, if groups choose to, as we did, include non-English speech in brackets to denote a different language, transcribers should be trained to make sure all open brackets are closed. We found this to be a common error in our transcripts. Further, if transcribers use a non-standard keyboard to transcribe non-English text, it is important that the brackets that enclose the text remain standard brackets so that computer programs can detect them. Alternately, if non-standard brackets are used, they should be well documented and included in any code or regular expressions used for analysis. Finally, because space breaks

are treated differently by computer programs than continuous text, transcribers should be trained to avoid using line breaks in their transcripts. One could get around this problem by first removing line breaks from the entire transcript, unless of course, line breaks are used to distinguish between different audio files.

Prior to beginning data transcription, groups may also want to consider how to indicate certain words and expressions. For example, how should truncated words such as *'bout* for *about* be treated? Different groups may have different approaches or may want to use special characters or code to indicate both. For example, in the CHAT transcription format commonly used by language development researchers (MacWhinney, 2000), truncations are coded with parentheses—for example, *(a)bout* for *about* or *(re)member* for *remember*—which allows the researcher the option to either merge truncations with full forms or keep them separate. Groups also may want to develop a code for different usages of words with multiple senses. For example, in our transcripts we wanted to distinguish between colloquial uses of the word *like* and the use of *like* as a verb, so we coded the colloquial use as *like** (though to avoid the use of punctuation, groups could also consider a code such as *likelike* to aid analysis). Finally, because programming languages like Python can sometimes recode characters with accents or diacritics as its unicode string, groups may want to be aware of this, and make sure transcribers are being consistent with their use of accents and diacritics.

Coding Category	Wave 1 ICC	Wave 2 ICC	Average ICC Across Waves
Discussing the EAR	.829	.925	.877
Discussing Aspects of Study	.726	.522	.624
Alone	.939	.945	.942
With One Person	.879	.896	.888
With Two or More People	.900	.925	.913
On the Phone	.959	.948	.954
Gender of Conversation Partner	.914	.795	.855
Speaking to Self	.796	.952	.874
Speaking to Known Person	.906	.924	.915
Speaking to Stranger	.787	.909	.848
Speaking to Child	.770	.917	.844
Speaking to Pet	.930	.964	.947
Radio/Music in Background	.930	.979	.955
Music Language	.866	.964	.915
Gaming	.963	.992	.978
TV/Video Language	.904	.872	.888
Computer/Texting	.892	.743	.818
Studying	.868	.853	.861
Eating	.841	.645	.743
Sports/Exercise	.981	.976	.979
Laughing	.909	.714	.812
Singing	.972	.931	.952
Mad/Arguing	.696	.560	.628
Apartment/Dorm/Other Residence	.986	.926	.956
Classroom	.934	.961	.948
Outdoors	.840	.882	.861
In Transit (Vehicle)	.714	.936	.825
In Transit (Other)	.952	.628	.790
Bar/Coffeeshop/Restaurant	.896	.917	.907
Shopping	.842	.943	.893
Other Public Place	.964	.871	.918
Participant Language 1	.708	.843	.776
Participant Language 2	.984	.625	.804
Participant Language Switching/Mixing	.895	.966	.931
Conversation Partner Language 1	.752	.646	.699
Conversation Partner Language Switching/Mixing	.848	.897	.873
Average ICC	.874	.858	.867

Supplementary Materials

Peer Review History

Download: https://collabra.scholasticahq.com/article/36310-using-the-electronically-activated-recorder-ear-to-capture-the-day-to-day-linguistic-experiences-of-young-adults/attachment/91821.docx?auth_token=5B8QsW5rwx_8PulZlo1M
