

**Best Practices for Electronically Activated Recorder (EAR) Research:
A Practical Guide to Coding and Processing EAR Data**

Deanna M. Kaplan^{1*}, Kelly E. Rentscher², Maximilian Lim¹, Ramon Reyes¹, Dylan Keating¹,
Jennifer Romero¹, Anisha Shah¹, Aaren D. Smith¹, Kylee A. York¹, Anne Milek^{1,3}, Allison M.
Tackman¹, Matthias R. Mehl^{1*}

¹ Department of Psychology, University of Arizona

² Cousins Center for Psychoneuroimmunology, Semel Institute for Neuroscience & Human
Behavior, University of California, Los Angeles

³ Department of Psychology, University of Zürich

* To whom correspondence should be addressed:

Deanna M. Kaplan (email: deannamkaplan@email.arizona.edu) or

Matthias R. Mehl (email: mehl@email.arizona.edu)

University of Arizona, Department of Psychology, Tucson, AZ 85721

Abstract

Since its introduction in 2001, the Electronically Activated Recorder (EAR) method has become an established and broadly used tool for the naturalistic observation of daily social behavior in clinical, health, personality, and social science research. Previous treatments of the method have focused primarily on its measurement approach (relative to other ecological assessment methods), research design considerations (e.g., sampling schemes, privacy considerations), and the properties of its data (i.e. reliability, validity, added measurement value). However, the evolved procedures and practices around arguably one of the most critical parts of EAR research – the coding process that converts the sampled raw ambient sounds into quantitative behavioral data for statistical analysis – have so far been largely communicated informally between EAR researchers. This article documents the “best practices” for processing EAR data that have been tested and refined in our research over the years. Our aim is to provide practical information on important topics such as the development of a coding system, the training and supervision of EAR coders, EAR data preparation and database optimization, the troubleshooting of common coding challenges, and coding considerations specific to diverse populations.

Keywords: ambulatory assessment, ecological momentary assessment, naturalistic observation, behavioral observation, smartphone sensing

Best Practices for Electronically Activated Recorder (EAR) Research:**A Practical Guide to Coding and Processing EAR Data**

The Electronically Activated Recorder (EAR) is an ecological assessment tool for the naturalistic observation of everyday social life (Mehl, 2017). The EAR is a digital audio recorder, currently available as an app for Android, formerly an iPhone app (“iEAR”), that can be downloaded onto smartphone devices and then worn by participants as they go about their normal daily lives. The app passively and intermittently samples brief snippets of ambient sounds from the wearer’s momentary environment. Sampling rates are customizable and may vary by study, but typical designs record 30 or 50 seconds of audio 3 to 10 times per hour (approximately 5-10% of the wearer’s day). Through high-frequency sampling of ambient soundbites from morning to night, the EAR provides “acoustic diaries” of individuals daily lives. These sound files can then be 1) behaviorally coded, yielding quantitative data about participants’ moment-to-moment environments, behaviors, and social interactions, and 2) transcribed, yielding natural daily language use data that can be processed with linguistic analysis software.

The EAR was first introduced to the scientific community in 2001 through an article in this journal (Mehl et al., 2001). Initially considered a somewhat esoteric “niche-method”, it has over the years methodologically matured into a psychometrically established and broadly used ecological assessment method. To date, the EAR method has been successfully employed, with good acceptance and compliance, in a range of healthy and clinical populations (Baddeley Pennebaker & Beevers, 2013; Brown, Tragesser, Tomko, Mehl & Trull, 2014; Hasselmo et al., 2018; Holleran, Whitehead, Schmader & Mehl, 2011; Minor et al. 2018; Robbins et al., 2018; Slatcher & Robles, 2012; Tobin et al., 2015) and in age groups ranging from early childhood to

old age (Alisic, Barrett, Bowles, Conroy & Mehl, 2015; Bollich et al., 2016; Demiray & Martin, 2018).

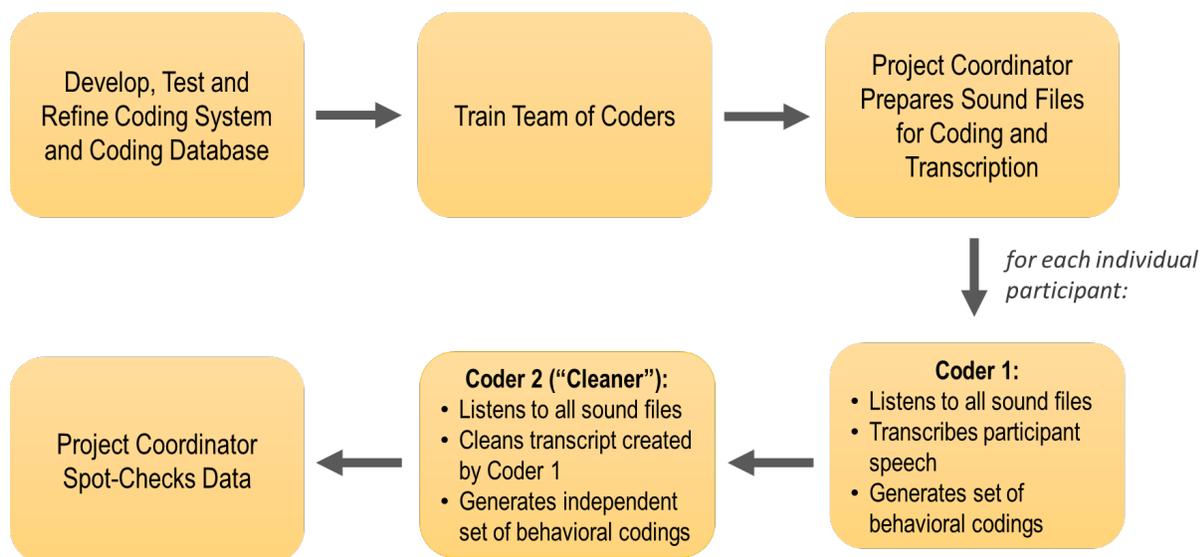
Existing treatments of the method have focused on conceptualizing the EAR within the broader “toolbox” of ambulatory assessment methods (Mehl & Robbins, 2013; Mehl, 2017), establishing a measurement rationale for it (Alisic et al., 2015; Mehl, Robbins, & Deters, 2013), and addressing important research design considerations including method acceptance and adherence (Mehl & Holleran, 2007; Manson & Robbins, 2017) and the protection of participant and bystander privacy (Robbins, 2017). Our goal for this article was to document the accrued experiences and evolved procedures and practices around the coding process – that is, the process of converting the sampled raw ambient sounds into quantitative behavioral data for statistical analysis. This is arguably one of the most critical aspects (and certainly the most time-consuming part) of EAR research, but as of yet has not been formally documented. In doing so, this article attempts to offer suggestions for “best practices” on the basis of what, over time, has proven useful in our research. A thorough documentation of what we have found to “work” with regards to processing the EAR audio data not only has the potential to save future EAR researchers time, it is also critical from an open and transparent science perspective to facilitate the reproducibility of EAR research (Nosek et al., 2015; Zwan, Etz, Lucas & Donnellan, 2018).

Over the last 20 years, our lab has tested and refined procedures for developing coding systems, practices for training and supervising teams of coders, and strategies for EAR data preparation and database management to increase the reliability and efficiency of projects. This paper describes the current “best practices” in our lab, which we recommend as default or “starting point” considerations for researchers. We anticipate that, as the method continues to be employed in new contexts with new populations and new kinds of research questions, the field’s

collective repertoire of best practices will continue to expand and evolve. Further, we want to stress that the “best practices” we offer here will necessarily require minor or major modifications and adjustments to accommodate individual cases. In the spirit of collaborative and evolving science, we therefore welcome researchers to share the tools, protocols, procedures, and practices that they have found helpful with populations and questions they have studied as common resources on the Open Science Framework’s EAR Repository at <https://osf.io/n2ufd/>.

Figure 1 provides a “bird’s eye view” of steps to the coding and processing of EAR data that are described in detail in this paper.

Figure 1. Steps to coding and processing EAR data



EAR Coding System Development

There are two primary approaches to extracting relevant information from EAR-derived audio recordings: psychological rating and behavioral coding approaches. The first involves extracting information at a *molar* or “broader” psychological level, in which expert raters listen to each participant’s sound file (or set of sound files) and rate the *degree* to which a relatively broad construct of interest is present. For example, a relationship scientist may choose to rate the

degree of expressed emotion or responsiveness that is captured in the participants' interactions with family members. In contrast, the second approach involves extracting information at a *molecular* or "smaller" behavioral level, in which expert coders listen to each sound file and make a binary coding that indicates the *presence* (or absence) of a comparatively narrow behavior of interest. Using the previous example, in this case, a relationship scientist could code the presence of an expression of affection or an anger outburst, in the captured conversations with the participants' family members.

In our research, we have primarily adopted a molecular approach to extracting information from EAR-derived audio recordings; however, we also direct interested readers to research by other investigators that has adopted a more molar approach. For example, Sun & Vazire (2019) used a 5-point scale to rate participants' personality dimensions (e.g., extraversion, agreeableness) over 1-hour intervals (i.e., each interval included 6-7 audio files) to coincide with participants' hourly ecological momentary assessments of their personality using the same scale. Of note, the authors explained that they increased the number of coders to six per participant after low initial inter-coder reliability with three coders per participant, and commented that some of the dimensions (e.g., neuroticism) were difficult to detect acoustically. Other investigators have rated the degree of specific emotions expressed (e.g., upset or happy; 5-point scale), maternal responsiveness (a composite comprised of expressed warmth, emotional support, and pride; 5-point scale), and the overall emotional tone of an interaction (ranging from negative to positive; 7-point scale) in each sound file with acceptable to excellent inter-coder agreement (Alisic et al., 2015, 2017; Farrell et al., 2019; Mangelsdorf, Mehl, Qiu, & Alisic, 2019; Tobin et al., 2015a; Tobin et al., 2015b).

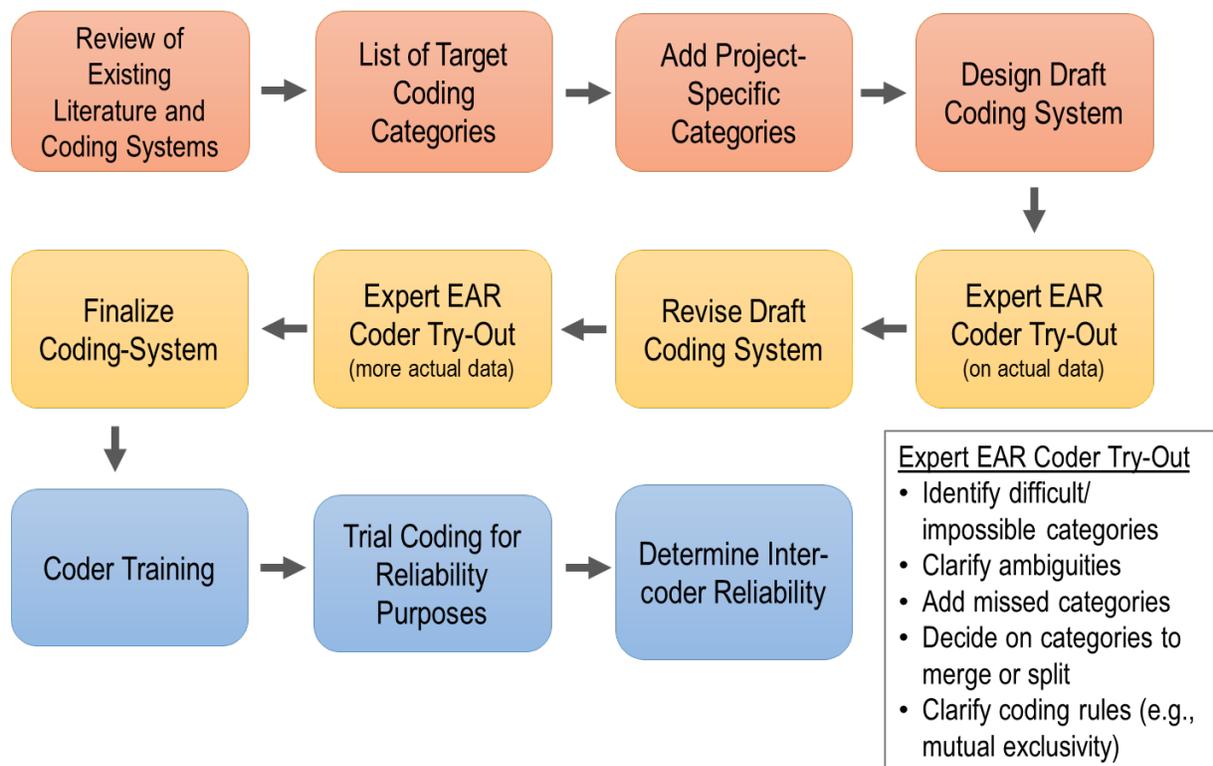
In our laboratory, trained coders listen to all of the available recordings for a participant

and code each sound file, consecutively, for the presence of specific behavior categories using a standardized coding system. This coding system was coined the Social Environment Coding of Sound Inventory (SECSI; Mehl et al., 2006; Mehl & Pennebaker, 2003). The basic SECSI consists of several core modules that were designed to capture acoustically detectable aspects of participants' social environments and interactions. The core modules include: (1) the person's current location (e.g., at home, outdoors, in transit); (2) activity (e.g., listening to music, watching TV, eating); interaction (e.g., alone, talking, on the phone); emotional expression (e.g., laughing, crying, sighing); interaction partner (e.g., friend, romantic partner, stranger); and conversation type (e.g., small talk, substantive, gossip). Coders typically transcribe and complete behavioral codings in parallel, basing behavioral codings on speech content as well as other contextual information (e.g., tone of voice, setting, speech and behavior of other individuals captured in the sound file).

Over the course of various projects, we have modified the basic SECSI coding system to include behavior categories (e.g., locations, activities, and interaction partners) that are unique to individual projects and their respective research questions. For instance, in a study of breast cancer survivors, we expanded the *interaction partner* module to capture interactions with participants' medical providers, and in a study of behavioral manifestations of meditation training, we expanded the *activity* module to include when participants were meditating. We have also designed new, project-specific modules to capture, for example, breast cancer survivors' interactions with their support networks (e.g., positive and negative received support; Robbins et al., 2014), and pro-social behaviors in the context of meditation training (gratitude, affection; Kaplan et al., 2018). In recent projects, we have further expanded the SECSI system to assess family-level environments and interactions (e.g., Alisic et al., 2015; Mascaró, Rentscher,

Hackett, Mehl, Rilling, 2017), which required significant modifications to the basic coding system in order to capture the complexities of daily family life. Examples of these and other coding systems are available by request and as part of the OSF EAR Repository. Based on these experiences, the remainder of this section outlines our process for developing EAR coding systems (Figure 2), with recommendations and considerations to meet the unique needs of each research project and population.

Figure 2. Schematic overview of important steps in developing an EAR coding system



Review existing literature and coding systems. The first step in designing a new EAR coding system is to conduct a thorough review of the literature to identify important constructs and existing coding systems related to the topic of interest. Although researchers may begin with familiarity of the literature on a given topic, we encourage a broad survey of the literature to identify a range of relevant constructs (e.g., those with prognostic value in previous research)

and coding systems (e.g., established laboratory-based ratings systems), as some may be better suited for acoustic-only detection than others. This step also includes brainstorming theoretically relevant behavior categories that may not have emerged during the review but would be important given the research project or population.

Create list of target coding categories. The second step involves creating a list of target coding categories by narrowing down the potential constructs and behaviors identified during the literature review. At this stage, it is important to note that laboratory-based coding systems often do not translate well when applied directly as EAR coding modules. This is in part because observational coding systems in the laboratory are often applied to interaction tasks that are of longer temporal duration than EAR-derived audio files (e.g., 10 minutes vs. 30 seconds) and rely on input from multiple channels (i.e., visual and audio). For instance, a relationship researcher who is interested in coding demand-withdraw interaction patterns may have difficulty capturing a complete, reciprocal interaction between two people within a 30- to 50-second timeframe. In this case, the researcher might consider coding instances of demand behavior, although withdraw behavior may be more challenging to detect with acoustic information alone. It is therefore essential that researchers consider whether constructs of interest—especially those that are more *molar* in nature—can be operationalized at the *molecular* level, as concrete instances of behavior that can be captured acoustically within the available timeframe. In addition, whereas laboratory interaction tasks are typically constrained to a specific type of interaction (e.g., a conflict or problem-solving discussion) and provide a clear context for coding behaviors of interest, these interactions may be more fluid in daily life, and therefore more challenging to detect and code. In this case, we and other researchers have found it helpful to include a broader category to help identify a context of interest (e.g., interaction with spouse, conflict), followed by a set of specific

behavior categories that apply to that context (e.g., demand behavior, criticism; Wang & Repetti, 2014, 2016). Another recommendation at this stage is to select behaviors that are maximally intuitive to observers; for example, most coders recognize an expression of gratitude when they hear it.

Add project-specific coding categories. The third step involves reviewing the SECSI core modules, and deciding whether to modify coding categories within existing modules and/or add project-specific coding modules. Given that many of our projects aim to investigate the psychological and social factors that facilitate adjustment to a stressful life event or experience, we typically include a category to capture conversations that are about the target event of study itself (e.g., cancer, an accident or physical injury). This allows us to assess the frequency and quality of conversations about the event or experience in participants' daily life, which we have found can have prognostic value (Robbins et al., 2018; Alisic et al., 2017). The total number of categories in the coding system is an important consideration at this stage. We have found that even expert EAR coders can become fatigued and miss categories if there are too many behaviors to code. Second, if an investigator has constraints on time or budget, it may be helpful to identify and focus on the top coding categories for a first pass through the data. It is always possible to code additional categories as a second pass at a future date, then with the "built-in efficiency" of being able to focus on a specific subset of audio files (e.g., only coding files in which participants are talking about divorce or interacting with a romantic partner).

Design draft coding system. The fourth step involves designing a draft coding system that includes a project-specific coding manual with definitions and rules for coding each behavior category. At this stage, it is important to consider the workflow of the EAR coders, and to think through the order of the coding categories to optimize efficiency and accuracy. We have

found it useful to order the coding modules based on complexity so that coders are presented with lower-level categories first (e.g., location, activities), followed by higher-level categories such as conversation type or specific interactions. This can improve efficiency as complex categories often require multiple reviews of an audio file. In addition, we have found that if the definition of a specific coding category requires more than a brief description, it is probably not intuitive and may present challenges for inter-coder reliability. Finally, we also consider whether or not groups of coding categories within a module should be designated as mutually exclusive (i.e. *only one category* within a module can be marked as present per sound file). For example, we typically force the *conversation type* module to be mutually exclusive, such that in any given sound file, the participant can be coded as engaged in small talk or a substantive conversation, but not both. We have found that coding modules in this way can help make the relative frequencies of these coding categories more intuitive and easier to interpret. Mutual exclusivity in coding categories can be forced through coding specifications in RedCap (see section on RedCap coding provided below). When categories are designated as mutually exclusive but coders perceive that more than one coding category is present (e.g., a sound file begins with smalltalk but subsequently transitions to substantive conversation for the majority of the soundfile), coders are instructed to code the category that best captures the dominant theme of the sound file (in this example, substantive conversation).

Expert EAR coder tests coding system. The fifth step involves enlisting an expert EAR coder to test the coding system on actual data. Our approach has typically been to have an experienced coder in our lab (1) listen to and code all of the available recordings for a couple of participants, and then (2) sample audio files from several additional participants at different times of the day in an effort to get a full range of potential environments and interactions. For

investigators that are new to EAR research, it may be helpful to train a naïve coder (see section on EAR coder training) who was not involved in the development of the coding system to test the system on actual data, and then compare the codings to those of a more experienced coder (i.e., the investigator or another person involved in the coding system development). During this process, we encourage the coder to keep detailed notes about the coding process (outlined next).

Revise draft coding system. The sixth step involves revising the draft coding system based on feedback from the expert EAR coder. At this stage, it is important to solicit feedback from the EAR coder about categories that were difficult (or impossible) to code, or any relevant categories that we missed but might want to add. It is also important to clarify any ambiguities with the coding manual definitions or the coding rules. During this process, researchers may notice that some of their theoretically derived EAR categories occur with low frequency (or not at all) in the subset of data used to test the coding system. In this case, researchers might consider dropping some of these categories; however, we recommend retaining categories that are central to the research question regardless of their base rates during the testing of the system, as they may provide information about the frequency of specific behaviors in daily life that may be counterintuitive and stimulate new research directions. For example, in a study of breast cancer survivors and their spouses, we were surprised to learn that cancer-related conversations comprised approximately 5% of the couples' conversations (Robbins et al., 2014, Robbins et al., 2018). During this step, it is also helpful to solicit coder feedback about the length and scope of the coding system. Although there is no “optimal” number of coding categories, very long coding systems may slow down the coding process or tax the working memories of coders, leaving infrequently used categories vulnerable to being forgotten. Based on this feedback and theoretical considerations, we then decide which categories to merge, split, or remove, in order to

optimize the length and scope of the coding system.

Expert EAR coder tests revised system. The seventh step involves testing the revised coding system on data from a couple of additional participants.

Finalize coding system. The eighth and final step involves revising and finalizing the coding system based on feedback from the expert EAR coder.

Special considerations for working with vulnerable populations. As many of our projects investigate the social environments and interactions of individuals coping with stressful life experiences, our research often includes participants who are psychologically vulnerable. Therefore, it has also been important that our coding systems include a category to assess the presence (or potential) of harm or abuse—either to the participant him/herself, or to others in the environment. This has been particularly important when conducting research with children and adolescents, and we recommend instructing coders to inform investigators immediately if there is any suspicion or concern of harm to self, harm to others, abuse, or neglect. In our lab, we intentionally set a low threshold for the definition of harm or abuse, in order to shift the burden of making a decision about whether an interaction meets the criteria for harm (and is therefore subject to reporting requirements) from the coder to the investigators. The safety of participants is an essential consideration in all EAR research, but is of particular importance when working with psychologically vulnerable populations, children and adolescents, and older adults.

The Human Side of EAR Research: Training and Supervising EAR Coders

After the development of a theoretically meaningful and practically feasible coding system, the single most important ingredient to the successful processing of EAR data is the recruitment, training, and retention of well-trained EAR coders. Coders have the critical responsibility of converting the qualitative and ethnographic nuance of EAR sound files into

quantitative data that can be statistically analyzed. In a typical EAR study, each coder will listen to several thousand sound files, while study investigators may only listen to a small number. Coders are the first-line researchers in any EAR study and come to know individual participants as well as audible trends in the dataset overall in a way that investigators simply cannot. EAR investigators rely on coders for insights about the data and the performance of the coding system. Thus, while the training of EAR coders is a non-trivial time commitment, it is time well-invested. In this section, we offer recommendations for the recruitment, training, and professional development of EAR coders. Importantly, different models of recruitment, training and development will work best in different labs. We describe the practices that we have found to be optimal for our own lab and encourage researchers to adapt these recommendations to fit their unique project-specific and lab-specific needs.

Recruitment and qualifications. Our lab typically recruits undergraduate student coders. We have historically welcomed new coders into the lab who have no or little prior research experience: given the unique nature of behavioral coding and the extensive training that EAR coding requires, we have not found prior research experience to be predictive of coder success. The completion of a Research Methods course is helpful, as is coursework in content areas relevant to the focus of the project. For example, in our lab's prior study of oncology populations, the perspective of pre-medical students was beneficial (Robbins, Lopez, Weihs, & Mehl, 2014); in our current and ongoing study of the role of family environments in adolescent suicidality, having coders with academic backgrounds in family studies has been valuable (though not essential) to the task of coding complex family environments. Bilingualism is also an asset for many projects (see *Multilingual Coding and Transcription*, below). We typically aim to recruit students in their Sophomore or Junior year of their undergraduate career; given the time

investment for coder training and the challenges that coder turnover can create in the middle of a project (e.g., “coder drift”), the longer coders can be retained, the better.

When we interview prospective coders, we have found that being upfront about the day-to-day reality of the (frequently humdrum) coding work is helpful for assuring a good fit between the student and the nature of the work. We typically advise students that while the job of eavesdropping on daily life may sound exciting and entertaining, the majority of daily life is quite uneventful—mundane activities of living such as eating, working, and watching television are boring to listen to and tedious to code. On the other hand, it is also important for prospective coders to know that emotionally charged, bizarre, and highly sensitive events are sometimes recorded. The EAR captures a representative sample of daily life, and when participants do not exercise the option to delete sensitive files (which they rarely do) this can include heated arguments and expressions of emotional distress, as well as activities that are of a private nature such as using the bathroom or having sex. Given that coders are the first to hear the sound files, the job thus requires a tolerance for the tedium of the mundane as well as comfort with hearing a broad spectrum of behaviors that are characteristic of the human condition.

Training. Our experience has been that a thorough training of EAR coders takes 6-8 weeks. We break training into three steps: coding/transcription training, a practice coding trial, and troubleshooting (guided by inter-coder reliabilities derived from the practice coding trial).

Coding/Transcription Training. To train new coders, we typically offer a workshop presented over a series of 2-3 meetings. We begin by presenting coders with a brief summary of how EAR research works, and what the added value is of using this methodology for the specific project. In order to avoid biasing the coding process, our practice has been to keep coders naïve to the specific research questions and hypotheses of the study, and to instead talk about the

purpose of the project in broader terms. When extra EAR devices are available in the lab, we often encourage new coders to experiment with wearing the EAR themselves for an evening or day and listening to the resulting soundbites of their own daily life. This provides coders with a first-person sense of what the EAR captures, as well as what the experience of participating in an EAR study (i.e., wearing a device that intermittently records ambient sounds) is like.

We next train coders in transcription. Rules for transcription will vary based upon the software used for post-processing of transcripts. Our lab has historically used Linguistic Inquiry and Word Count (LIWC), and so the current transcription training we provide is based on the LIWC 2015 software manual (Pennebaker, Boyd, Jordan & Blackburn, 2015). Given the specificity of transcription rules, we have found it helpful to incorporate interactive practice into this aspect of training. We typically provide coders with a few sample transcripts that contain errors, and ask the group to collaboratively identify and correct the transcription errors.

The final (and most time-consuming) part of coding and transcription training is the review of the coding manual. Our practice has been to review the coding manual in its entirety out loud as a group. When sound files for the project are already available, we play examples of each coding category to illustrate the range of behaviors that each coding category is intended to encompass. We conclude coder training by coding a handful of practice sound files as a group.

Trial of Practice Participants. It is not uncommon for new EAR coders to report that coder training feels a bit like “drinking out of a fire hose” -- the number of transcription and coding rules that coders are asked to learn is considerable, and questions at this stage tend to be numerous. We have therefore found giving coders the opportunity to practice applying the transcription and coding system to actual participant sound files to be an essential component of training. Typically, we select a subset of 150 consecutive sound files for two participants who

are ideally psychologically and behaviorally quite different (for a total of 300 sound files). This provides enough variability that most coding categories, even those with low base rates, have decent odds of surfacing at least once. Coders record questions that come up during this process, and we meet regularly (once a week) as a group to discuss these questions. Resolving questions about the coding system as a group at this stage minimizes discrepancies in understanding between coders, which is critical for promoting adequate inter-coder reliability later on.

Troubleshooting. As a last step, we compute intercoder reliabilities for all variables across both practice participants to examine inter-coder agreement, and subsequently hold a meeting with our coders to review this information and discuss categories that did not achieve excellent agreement (i.e., Intraclass correlation coefficients [ICCs] less than or equal to .70). Our practice has been to approach these meetings by explaining to our coders that low agreement indicates to us that we have room for improvement in our coding system. For example, low agreement may indicate that a category needs to be rendered more concrete). Coders are only able to capture what the coding system allows; thus, low agreement is a result of a suboptimal coding system rather than a lack of the coders' ability. Therefore, we typically begin by asking our coders, "What was difficult about this category – why do you think we might have low agreement?" Often, coders will have identified unforeseen ambiguities in the coding system that can then be clarified through discussion. Asking coders whether it was difficult to distinguish the coding category in question from other categories can be an important follow-up question; if coders have difficulty discriminating between two categories, this reduces the reliability of both categories, and is often easily rectified by refining the category definitions and exemplars. Another useful follow-up question can be, "Who remembers using this category at least once? At least 5 times? At least 10 times?" Low base rates of the behavior in the training set is a common

culprit for low intercoder reliability at the training stage and does not necessarily mean that there is a problem with the coding category. Given that low base rates constrain inter-coder agreement statistics, we find it helpful to have base rate statistics on hand for these meetings to facilitate the interpretation of inter-coder agreement in the context of base rates. Many behaviors of interest to daily life that typically exhibit low base rates at the training stage (e.g., conflict) occur frequently enough over the life of a sufficiently powered study that intercoder reliability is minimally impacted.

Coder development and coding ambiguous situations. Coders have consistently reported to us that the most difficult aspect of coding is acquiring confidence in one's coding decision-making. This of course comes with time and practice, but there are things that EAR researchers can do to help coders build self-efficacy. One is to convey that the research team expects there to be times when the answer is simply ambiguous or unclear. In designing a coding system, it is impossible to foresee all possible combinations of human interaction and behavior that may surface in the data, and EAR research accounts for this through its high sampling rate or, in other words, an "oversampling" of occasions. In a study with more than 50,000 sound files, the coding of a small number of ambiguous sound files is not going to substantially affect any statistic. As a rule, we therefore encourage coders to try not to spend too much time on any given sound file. Occasionally it can be helpful to listen to a sound file twice or, at most, three times (e.g., when there is a lot of background noise or the participant is speaking quietly), but listening to a sound file any more than this is rarely, if ever, warranted. When ambiguous coding situations arise, we advise coders to take their best guess, consulting with their project coordinator or each other if needed. We also encourage coders to bring such questions up during the weekly lab meeting, in which we discuss coding questions and ambiguous coding situations.

Importantly, “taking your best guess” involves reasoned decision-making, not random choice. As questions arise, we have therefore found it beneficial to help coders in developing a systematic decision-making process. The most common type of question that comes up, especially among newer coders, relates to distinguishing between categories – for example, “Is this small-talk or a practical conversation?” In these cases, we typically ask our coders to review the manual definitions for each category in question and to then make a case for each category (“Which way are you leaning and why? What would be the argument for the other category?”) If the question is presented at a lab meeting, our practice has been to involve all coders in arguing for and against the coding categories in question. While the impact of the coding decision about the sound file in question is statistically negligible, the process of practicing reasoned coding decision-making has benefits throughout the life of the project and can also yield important insights about how the coding system is performing.

Coder integrity. In any EAR study that involves double-coding of participants, it is critical that both sets of codings remain independent. Thus, while we encourage coders to ask each other questions about ambiguous coding decisions, we also ask them to avoid asking questions to the other coder for their participant. Sometimes this is unavoidable; for example, if only two or three coders are working on the project. In these cases, we advise that *general coding questions* are appropriate (e.g., “should I transcribe ‘yo’ as a non-fluency?”; “does babysitting for a younger sibling count as a chore?”) but *specific questions* must be avoided (e.g., “is the participant dating that Antonio guy that keeps showing up in the first week of files?” or any question that requires directly quoting from the sound file to ask).

Efficient and effective coding: optimizing the workflow. Coding and transcription are, by far, the most time-consuming part of any EAR project. The amount of time it takes to code a

single participant can be estimated by summing the net audio time and multiplying it by a factor of 5 or 6 (to account for time spent listening to the sound file, transcribing, and generating the codings). While the time-consuming nature of EAR research is to a large extent unavoidable, we offer recommendations for equipment, documentation and lab practices that we have found to maximize efficiency in our own lab.

Equipment. Two core pieces of equipment are necessary for EAR coding: headphones and transcription software. There are a number of headphones on the market, and the features we have found to be most important are high sound quality, comfortable ear pads for prolonged use, adjustable ear pads for coders who wear eyeglasses, and noise cancellation (for use in lab spaces where coders will be seated in close proximity to one another). Given that transcription is the most time-consuming component of processing EAR data, we additionally recommend the use of a transcription foot pedal in conjunction with transcription software to maximize the efficiency of this process. This equipment allows coders to quickly rewind and fast forward within the sound file using their feet while simultaneously transcribing with their hands.

EAR project documents. We have found a “Project Status Sheet” to be instrumental to the organization of an EAR study. In our lab, this has taken the form of an Excel spreadsheet which contains a list of all participants available for coding. To avoid any incidental duplication of work, we ask our coders to use this spreadsheet to “sign up” to code and clean participants. Our coders also use this document to record other information about the EAR data for each participant, such as the total number of sound files, the number of un-coded sound files (e.g., if only sound files up to a certain date and time are being coded), the coders’ estimation of compliance with wearing the EAR device, and the presence of foreign language in the data.

We additionally ask coders to complete a “Comments about the Participant” document

for each participant coded. This a place for coders to document noteworthy overall (rather than sound-file specific) comments, observations, and issues or uncertainties. This includes notes such as “participant sounds similar to his younger brother but has a slightly deeper voice”; “most of the participant’s verbalizations and social interactions occur through Playstation”; “the participant has divorced parents and goes back and forth between their two homes.” Because these comments are general in nature, we permit the second coder for a given participant to read the first coders’ comments, and this can often expedite the second set of codings by reducing the amount of time that the second coder must spend on speaker identification. In addition, later on during data analysis, these documents can provide useful information for outlier analyses.

We have also found the maintenance of an “EAR Project Wiki” to be essential. This is a living document updated throughout the life of the project that contains a record of how all coding-related questions (e.g., “should babysitting a younger sibling be coded as ‘doing chores’”?) have been resolved. This assures that, should a similar question arise in the future, coding decisions are consistent throughout the life of the project.

Many EAR projects may also benefit from EAR Event Diaries and Participant Readme Files. EAR Event Diaries are self-reported records of daily activities completed by participants during the duration of wearing the EAR device. Participant Readme Files are short text documents prepared by the data collection team for the coding team to provide contextual information about the participants’ life. This might include participants’ age, gender, with whom they reside (e.g., which parent in the case of custody arrangements), their profession, and diagnosis and treatment plan (for medical populations). In studies with complex family environments or studies that feature activities for which coders may not readily recognize the sounds (e.g., receiving dialysis), providing this information can help improve coding accuracy.

EAR lab practices and minimizing “coder drift.” The best means of expediting an EAR study is to retain coders for as long as possible – ideally, over the full life of the project. The efficiency of EAR coding dramatically increases with coder experience, and training a new cohort of coders can pause progress on a project for months. Moreover, “coder drift” can hurt the reliability and validity of coding categories when cohorts of coders differ in their collective understanding of categories.

The biggest obstacle to retaining coders is burnout. As has been emphasized, much of daily life is mundane and unremarkable, and coding can therefore be an extremely dull and tedious experience. We therefore recommend that coders arrange their schedules in order to avoid coding for more than 3 hours at a time. Some patient populations can also leave EAR coders vulnerable to issues similar to clinician-burnout. Healthcare professionals spend the duration of an office visit with any given patient; in contrast, EAR coders follow participants for many days or weeks of their actual daily lives. In this sense, coders get to know participants in a more personal and intimate way than their healthcare providers do, and it is not uncommon for coders to report feeling a great deal of empathy for participants. This can be a fatiguing and even distressing experience in participant populations whose daily lives are characterized by a great deal of psychological and/or physical suffering. We have found weekly coder meetings to be instrumental to minimizing both sources of burnout – tedium on one hand, and empathic fatigue on the other. In addition to providing opportunities to discuss and resolve questions about the coding system, these meetings also provide coders with the opportunity to share with each other and the research team about what they are hearing. Coding is by necessity a solitary activity, and over the years our coders have told us that maximizing opportunities for social connectedness is critical to preventing burnout and maintaining morale.

It is often not logistically possible to retain the same cohort of coders over the life of a project. When this is the case, we have found it helpful to begin training a new cohort of coders before the outgoing cohort departs. We solicit the assistance of the outgoing group of coders in training the new cohort, and ask both old and new coders to be present for all coder meetings. This helps assure that the collective knowledge of the original coder group is shared with the new coder group, minimizing any coding continuity issues that can arise with “coder drift.”

Finally, and maybe most importantly, a straightforward way to maximize coder retention and satisfaction is to offer research assistants hourly pay. It is certainly possible to run EAR projects with students who code the sound files for research course credit, and we do so in cases where we do not have grant funding for a study. However, after one or two semesters, even the most engaged students feel they have reached a learning plateau. Paying coders allows them to stay committed to the project, prioritize their coder position over jobs they may otherwise need in order to pay the bills, and instills in them a sense of their critical role on the research team as the “ethnographic” researcher on the ground. Practically, in our experience, it works best to have coders start as research assistants for course credit for one or two semesters. However, after the second semester, most students transition to coding for hourly pay.

EAR Coding Challenges and Potential Solutions

At its core, behavioral coding is a process of applying human intuition in an a priori and methodologically rigorous way. The intuitive component is particularly important for troubleshooting – for example, identifying when something doesn’t quite “sound right” and having awareness about how our individual backgrounds color how we interpret the content of the files. We offer some guidance for three of the most common challenges that require troubleshooting in EAR research: speaker identification, multilingual coding and transcription,

and diversity and multiculturalism concerns.

Speaker identification. The first task a coder has when beginning a new participant is identifying the participants' voice. This is fairly straightforward when the first few sound files capture the participant talking in a relatively quiet environment (e.g., the car ride home from the research lab), or when participants talk about things that clearly identify them as the participant (e.g., participating in an EAR study, having a diagnosis that identifies them as the target participant). The task of speaker identification can be much more difficult when the participant speaks little, lives with many other people at home, or spends a lot of time in noisy, public environments. We encourage coders to listen to several consecutive sound files without coding until they feel confident that they know which voice belongs to their participant. Participant Readme files with contextual information about the participant (e.g., age, gender, who else lives at home with the participant) are a helpful speaker identification aid for coders. When feasible, it also helps to ask a member of the data collection team who has had contact with the participant and can easily identify their voice to "pre-listen" to a few sound files in order to identify a target sound file that contains a clear sample of the participants' voice. This can then be indicated to the coder in the Readme file for the participant.

To more effectively address the challenges associated with speaker identification, the newest version of the EAR app also contains a Baseline Speech Sample functionality that allows the research team to record a sample of the participant's speech to aid in speaker identification. A pre-recorded speech sample provides coders with a target voice for reference as they begin coding, reducing the amount of time spent on identifying the participant. We recommend having participants read a "poem" of 30 fixed selected sentences containing all English phonemes. The poem was assembled from the TIMIT Acoustic-Phonetic Continuous Speech Corpus with the

explicit goal of facilitating speaker recognition (Larcher, Lee, Ma & Li, 2014).

Multilingual coding and transcription. When participants speak a language other than English (e.g., at home), our practice has been to code and transcribe as much as we can, being mindful of the extent to which we are adding valid “signal” as opposed to “noise” to the data set. Multilingual coders are an asset to any EAR lab, and whenever possible, we assign multilingual participants to coders who are fluent in the language(s) spoken by the participant. When this is not possible, we ask the coder to make a judgment about whether monolingual coding and transcription is sufficient to yield representative information about the participant’s daily life. When the coder believes that it will be – for instance, when the participant speaks a foreign language a small minority of the time or to one person in their life only – our practice has been to code and transcribe as best as we can, omitting transcripts for sound files in which the participant is speaking a foreign language. If the participant speaks a foreign language the majority of the time or frequently in conversations with persons who are essential to the research aims of the study (e.g., the participant mostly speaks in Arabic to her spouse in a study of marital satisfaction), it may be best to suspend coding or transcription of the EAR data unless it can be completed by a multilingual coder, as the data may otherwise have critically limited validity.

When it is possible to match the languages spoken by the participant to a multilingual coder, our practice has been to ask the coder to translate all transcripts into English. This makes it possible to include language data from the participant in any analyses that will be run from the transcript data, but also introduces methodological limitations that are important to consider. First, many languages are spoken with different dialects around the world; for example, Mexican Spanish differs somewhat from Puerto Rican Spanish. When the coder speaks a different dialect of the language than the participant, they may need to listen to sound files more than once in

order to understand the participants' speech, increasing the time that it takes to code and transcribe the participant. Secondly, some utterances may not have direct English translations. This applies to specific words and phrases as well as to broader idiosyncrasies of speech (e.g., the English word "excited" does not have an equivalent translation in German that conveys the specific level of valence and arousal associated with the English meaning of this emotion). Bilingual coders should therefore be instructed to transcribe as *meaningfully* as possible rather than as *literally* as possible; after all, transcript data are (at least in our research) ultimately analyzed via bag-of-word-based computerized text-analysis, and consequently the psychological validity of the transcript is of greater importance than word-for-word accuracy. Literal translations may mis-translate idioms, and coders may sometimes need to cut out or add words to the participants' speech in order to convey the intended meaning. This task comes easily to native and fluent speakers of the language in question, but is more difficult for coders who only have an academic understanding of the language and may not be as familiar with the vernacular.

Finally, it is important to keep in mind that participants who speak multiple languages may use English differently than monolingual participants. One of our bilingual coders described coding an adolescent participant from a bilingual (Spanish/English) family. The participants' brother spoke to the participant in both Spanish and English, and when speaking in English, referred to their mother as "my" mom (e.g., "Can you tell my mom to come here?") While this may have led some to question whether the participant and her brother have the same mother or not, our Spanish-speaking coder recognized that this was an artifact of the brother's bilingualism; in Spanish, the word for "mom" directly translates to "my mom" in English. From a data analytic perspective, it is important to keep track of multilingualism and associated features of speech identified by coders. In this case, for example, the brothers' use of possessive pronouns – a

common target for linguistic analyses – may be skewed by his bilingualism.

Cultural considerations in EAR coding. While the presence of bilingualism makes the role of cultural factors more conspicuous, they are always salient in EAR research. Ultimately, behavior and social interactions – the bedrock of EAR research – are culturally constrained variables. Each participant has cultural and group identities that carry with them norms and structures (Markus & Kitayama, 2010). Some of these may be readily apparent in sound files (e.g., a Jewish family attends synagogue together) while others may be more difficult to recognize without in-group knowledge (e.g., a Jewish family braids challah together). Each coder listens to sound files through the lens of their own group identities, norms and structures.

When there is a mismatch between the cultural lens of the coder and the cultural context of the participant, this can undermine the validity of the codings. For example, emotional expressivity (e.g., yelling, effusive affection) has culturally relative norms. A coder from a social context in which expressing emotion with high arousal is the norm may have difficulty recognizing expressions of anger or endearment from a participant whose social context promotes lower arousal expressions of these emotions. Family systems are also culturally embedded, and familial dynamics are accompanied by their own norms and structures. In some sense, each family unit represents its own cultural microcosm; consider the earlier example of Emily, whose father was initially mistaken for her brother because they had a relationship that, through the sociocultural lens of our coder, sounded much more like sibling rivalry than a “regular” parent-child dynamic. Coders may also have different individual thresholds for when an emotional conversation becomes a “conflict” or a “fight.”

The best means of promoting cultural competency in an EAR lab is to recruit coders from diverse backgrounds who can help to identify cultural blindspots in the application of the coding

system. This is yet another way in which regular coder meetings are helpful, as these meetings provide a designated time for coders to consider and discuss how their respective backgrounds and experiences may impact how they are using coding categories. Perhaps especially in EAR research, it is a good practice to encourage all team members to reflect on their own cultural contexts and biases, and to consider how these may bear upon their role within the project.

EAR Data Preparation, Database Management and RedCap Coding

For the first almost two decades of doing EAR research, our team used a simple spreadsheet (e.g., Microsoft Excel) for the coding of the EAR data. An EAR coding template was built that, within a single spreadsheet, listed each coding variable in a given coding system (e.g., is the participant talking? Is the TV on?) as a column and had “placeholder space” for each recorded sound file in the rows. This EAR coding spreadsheet template was then copied for each participant and pre-filled with the participant-specific EAR recoding information (e.g., recording file names, dates, and times). After that, it was handed to the coders for coding. Going through all of a participants’ recordings, sound file by sound file, coders would then make binary codings by designating a “1” in the appropriate cells to indicate the presence of a given behavior for each sound file (the absence of a behavior was indirectly indicated by skipping over a category; at the end of a project all empty cells were filled with zeros).

This approach was simple and low-cost on the front-end (e.g., preparation and training). On the back-end, though, at the end of a project when all coding was done, the project manager would be tasked with manually “stacking” all of the participants’ individual coding spreadsheets into one overall, merged database. This process was time-consuming (“stacking” several hundred spreadsheets, two for each participant with implemented double-codings), vulnerable to human error (e.g., shifting of columns, accidental omission of a section of rows), and for large studies

routinely led to the merged database exceeding Microsoft Excel limits in the number of rows. Finally, after the stacking, the project coordinator needed to thoroughly check tens of thousands of lines for human coding errors (e.g., accidental extra key strokes such as “11”, or logical errors such as failing to mark “participant talking” as “1” when a verbatim transcript was present).

To address these issues, we recently brought EAR coding “into the 21st century” by migrating it to a database approach. While our Excel EAR coding templates continue to be available on the OSF EAR Repository (<https://osf.io/n2ufd/>), we now recommend Research Electronic Data Capture (REDCap; Harris et al., 2009) for EAR coding. REDCap is a free (for non-profit user) and secure web application for managing databases. REDCap was developed at Vanderbilt University in 2004 as a secure online data collection tool that met HIPAA compliance standards. Since then, it has grown impressively and the REDCap Consortium now counts more than 3,000 participating and contributing institutions from more than 120 countries, with more than half-a-million projects, and more than 800,000 users (<https://projectredcap.org/>).

At the front end, REDCap provides an intuitive (though somewhat basic), web-based user interface which can be used for setting up the EAR coding mask (or “template”). The interface provides automatic version control (the interface tracks the author of all changes and additions to the database) and allows for parallel working (multiple coders can access and code data simultaneously). At the back end, it hosts a secure database system with user account management, version control, and data export possibilities. Its server-based architecture makes it a safe research tool from a data storage and back-up perspective, and its compliance with HIPAA regulations is critical for patient safety and, more broadly, EAR data confidentiality. For collaborative, multi-site studies, investigators can readily access (upon permission) the coded EAR data via their REDCap user account, and, for example, combine them with different data

sources collected at other sites. Finally, REDCap is easy to use in its basic features and requires no specialized computer knowledge; setting up an EAR coding data entry mask is no more difficult than designing a web-based questionnaire using a standard online surveying tool. Below, we outline considerations that we have found helpful in creating coding templates for REDCap.

Optimizing REDCap for EAR coding. We recommend the *Text Box* option for all text information that will be entered (e.g., the transcript, the coder's comments about the sound file), and the *Checkboxes (Multiple Answers)* or *Multiple Choice – Radio Buttons* for all binary coding categories. Checkboxes allow coders to select several options within the coding category that may apply, whereas radio buttons force a single answer among the available options. For purposes of EAR coding, both of these options are more efficient for coder use than the other field types available through REDCap (e.g., dropdown menus).

Once all coding categories have been added, the entry mask will look like a web-based survey that can be completed for each EAR sound file. Ideally, coders should be able to code by scrolling down the page a single time; a coding system that requires coders to scroll up and down the page many times will reduce coding efficiency. Ordering categories in an intuitive way is therefore important when programming the survey. We recommend beginning the instrument with a field that coders can use to upload the sound file into REDCap so that it is directly associated with the database. For this, we developed customized back-end code that auto-extracts the following information directly from the sound file name and writes them into the respective fields: participant ID, file start date, file start time, sound file number, and coder number. In addition, this feature automatically populates the first coder's transcripts for the second coder to edit and clean. This code is available upon request from the corresponding author.

Next, we include a text box for entering the transcript followed by coding categories that would “rule out” any further coding of the sound file (e.g., the presence of a foreign language the lab cannot handle, or the participant being deemed non-compliant or sleeping). We then list higher frequency categories (e.g., the participants’ current location, who the participant is with, whether they are talking, to whom they are talking and what they are talking about). This is followed by lower frequency categories (e.g., activities the participant may be engaged in or aspects of the participants’ momentary social environment). At the end of the form, we provide a text box for any general coder comments about the sound file, and additional check boxes for important meta-information about the file (e.g., presence of identifiable personal information, to be able to delete the segment or file). The last field of the form asks the coder to indicate the Form Status for the record, and we provide three options for this: *Complete* (which indicates the coding is complete and no further actions are needed), *Incomplete* (which indicates that the coder has a question about the sound file that needs to be reviewed with a project coordinator), and *Unverified* (which indicates an erroneous record, such as a duplicate record or one entered accidentally which needs to be later deleted by the project coordinator).

We highly recommend the use of branching logic throughout the EAR coding instrument so that only relevant categories are visible to the coder. For example, branching logic can be used to only provide options for coding who the participant is talking to if and only if the coder indicates that the participant is talking. Minimizing the number of categories that coders must unnecessarily scroll past can cumulatively save a good deal of time.

Lastly, a filtered “coder report” can be created in REDCap for each coder. A coder report is a link in REDCap that coders can click on to view their progress on the participant they are currently coding, as well as easily navigate to any records that they later determine they need to

revise or recode. This page can also be used by a project manager for spot-checking.

Conclusion

Our goal in writing this article was to complement existing methodological resources on the EAR method by, for the first time, formally documenting the evolved procedures and practices around one of the most critical parts of EAR research – the coding process. We guided our decisions regarding which information to include by what questions researchers interested in the method have asked us over the years and what challenges around its use have been reported to us. We settled on the term “best practices” although, of course, it is clear that what has worked in our lab with our projects is not necessarily what works for researchers in other settings and with different projects. For example, in our lab, we try to employ double coding of all EAR sound files where possible to account for the fact that reliably inferring behavior from ambient sound, particularly for complex psychological constructs, can be challenging. Double coding may not be feasible for all researchers and all projects and alternative ways exist to ensure reliability (e.g., there is ultimately a psychometric trade-off between double-coding sound files at a lower sampling rate or single-coding sound files at a higher rate).

As we mentioned at the outset, it is our hope that, as researchers construct their experiences with the method, the collective knowledge around its use will broaden and evolve and will ultimately render what we document here incomplete or outdated. Further, it is likely that computational advances in behavioral signal processing (Narayanan & Georgiou, 2013) will, in the future, render the coding process more efficient (e.g., through automatic identification of sound files that are silent or contain speech) and that some aspects of EAR coding may ultimately become automated (Dubey, Mehl, & Mankodiya, 2016; Yordanova, Demiray, Mehl & Martin, 2019). Until then, we hope that the documentation of what we have learned over two

decades of EAR research will help save researchers time and frustration and make a highly labor intensive method more accessible.

Role of Funding Source

This research was supported by National Institutes of Health R03CA137975, R21HD078778, 3R01AT004698, 5R01AT004698, R01HD069498, R01MH105379, and R01MH108641.

Author Contributions

D.M.K. and M.R.M. developed the paper concept. D.M.K. and M.R.M. drafted the manuscript with assistance from K.E.R., M.L. R. R., D.K., J.R., A.S., A.D.S., K.A.Y., A.M., & A.M.T. All authors approved the final version of the manuscript for submission.

Acknowledgements

We thank all past and current members of the Naturalistic Observation of Social Interactions (NOSI) Lab for their contributions to developing and refining the practices outlined in this paper. In particular, we thank Megan Robbins, Fenne große Deters and Angela Carey for their contributions. We also thank our current team of EAR coders, Valeria Chavez, Darrian Corey, Julie Drake, Chris Griffith, Elieana Herrera, Skyler Kopit, Paloma Sainz and Colin Tidwell.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Open Practices

All materials are publicly available at the Open Science Framework's EAR Repository (<https://osf.io/n2ufd/>).

References

- Alisic, E., Barrett, A., Bowles, P., Conroy, R., & Mehl, M. R. (2015). Topical review: Families coping with child trauma: A naturalistic observation methodology. *Journal of Pediatric Psychology, 41*(1), 117-127.
- Alisic, E., Gunaratnam, S., Barrett, A., Conroy, R., Jowett, H., Bressan, S., ... & Mehl, M. R. (2017). Injury talk: spontaneous parent-child conversations in the aftermath of a potentially traumatic event. *Evidence-based mental health, 20*(4), e19-e20.
- Baddeley, J. L., Pennebaker, J. W., & Beevers, C. G. (2013). Everyday social behavior during a major depressive episode. *Social Psychological and Personality Science, 4*(4), 445-452.
- 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.
- Demiray, B., Mehl, M. R., & Martin, M. (2018). Conversational Time Travel: Evidence of a Retrospective Bias in Real Life Conversations. *Frontiers in Psychology, 9*, 2160.
- Dubey, H., Mehl, M. R., & Mankodiya, K. (2016, June). Bigear: Inferring the ambient and emotional correlates from smartphone-based acoustic big data. In *Connected Health: Applications, Systems and Engineering Technologies (CHASE), 2016 IEEE First International Conference on* (pp. 78-83). IEEE.
- Farrell, A. K., Slatcher, R. B., Tobin, E. T., Imami, L., Wildman, D. E., Luca, F., & Zilioli, S. (2018). Socioeconomic status, family negative emotional climate, and anti-inflammatory gene expression among youth with asthma. *Psychoneuroendocrinology, 91*, 62-67.
doi:10.1016/j.psyneuen.2018.02.011
- Narayanan, S., & Georgiou, P. G. (2013). Behavioral signal processing: Deriving human

- behavioral informatics from speech and language. *Proceedings of the IEEE*, *101*(5), 1203-1233.
- Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, *42*(2), 377-381.
- Hasselmo, K., Mehl, M. R., Tackman, A. M., Carey, A. L., Wertheimer, A. M., Stowe, R. P., & Sbarra, D. A. (2018). Objectively Measured Social Integration Is Associated With an Immune Risk Phenotype Following Marital Separation. *Annals of Behavioral Medicine*, *52*, 130-145.
- Holleran, S. E., Whitehead, J., Schmader, T., & Mehl, M. R. (2011). Talking shop and shooting the breeze: A study of workplace conversation and job disengagement among STEM faculty. *Social Psychological and Personality Science*, *2*(1), 65-71.
- Kaplan, D. M., Raison, C. L., Milek, A., Tackman, A. M., Pace, T., & Mehl, M. R. (2018). Dispositional mindfulness in daily life: A naturalistic observation study. *PLoS One*, *13*(11), e0206029. doi.org/10.1371/journal.pone.0206029
- Mangelsdorf, S. N., Mehl, M. R., Qiu, J., & Alisic, E. (2019). How do mothers and fathers interact with their children after an injury? Exploring the role of parental acute stress, optimism, and self-efficacy. *Journal of Pediatric Psychology*, *44*(3), 311-322. doi:10.1093/jpepsy/jsy107
- 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.

- Mascaro, J. S., Rentscher, K. E., Hackett, P. D., Mehl, M. R., & Rilling, J. K. (2017). Child gender influences paternal behavior, language, and brain function. *Behavioral neuroscience, 131*(3), 262-273.
- 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.
- 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.
- 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.
- 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.
- 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.
- Mehl, M. R. & Robbins, M. L. (2012). Naturalistic observation sampling: The Electronically Activated Recorder (EAR). In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life*. New York, NY: Guilford Press.
- Mehl, M. R., Robbins, M. L., & große Deters, F. (2012). Naturalistic observation of health-

- relevant social processes: the electronically activated recorder (EAR) methodology in psychosomatics. *Psychosomatic Medicine*, 74(4), 410-417.
- 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.
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... & Contestabile, M. (2015). Promoting an open research culture. *Science*, 348, 1422-1425.
- 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.
- 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.
- Robbins, M. L., Karan, A., Lopez, A. M., & Weihs, K. (2018). Naturalistically observing non-cancer conversations among couples coping with breast cancer. *Psycho-Oncology*, 27(9), 2206-2213.
- Slatcher, R. B., & Robles, T. F. (2012). Preschoolers' everyday conflict at home and diurnal cortisol patterns. *Health Psychology*, 31(6), 834-838.
- Sun, J., & Vazire, S. (2019). Do people know what they're like in the moment?. *Psychological Science*, 30(3), 405-414. doi:10.1177/0956797618818476
- Tobin, E. T., Kane, H. S., Saleh, D. J., Naar-King, S., Poowuttikul, P., Secord, E., Pierantoni, W., Simon, V. & Slatcher, R. B. (2015a). Naturalistically observed conflict and youth asthma symptoms. *Health Psychology*, 34(6), 622-631.

doi:<http://dx.doi.org/10.1037/hea0000138>.

- Tobin, E. T., Kane, H. S., Saleh, D. J., Wildman, D. E., Breen, E. C., Secord, E., & Slatcher, R. B. (2015b). Asthma-related immune responses in youth with asthma: Associations with maternal responsiveness and expressions of positive and negative affect in daily life. *Psychosomatic Medicine*, *77*(8), 892-902. doi:10.1097/PSY.0000000000000236
- Tomko, R. L., Brown, W. C., Tragesser, S. L., Wood, P. K., Mehl, M. R., & Trull, T. J. (2014). Social context of anger in borderline personality disorder and depressive disorders: Findings from a naturalistic observation study. *Journal of personality disorders*, *28*(3), 434-448.
- Wang, S. W., & Repetti, R. L. (2014). Psychological well-being and job stress predict marital support interactions: A naturalistic observational study of dual-earner couples in their homes. *Journal of Personality and Social Psychology*, *107*(5), 864-878.
doi:<http://dx.doi.org/10.1037/a0037869>
- Wang, S. W., & Repetti, R. L. (2016). Who gives to whom? Testing the support gap hypothesis with naturalistic observations of couple interactions. *Journal of Family Psychology*, *30*(4), 492-502. doi:10.1037/fam0000196
- Yordanova, C. Y., Demiray, B., Mehl, M. R., & Martin, M. (2019). Automatic detection of everyday social behaviours and environments from verbatim transcripts of daily conversations. *IEEE PerCom 19*.
- Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, *41*. doi.org/10.1017/S0140525X17001972