Usability of a Hands-Free Voice Input Interface for Ecological Momentary Assessment

Rebecca Adaimi
Dept. of Electrical and Computer Eng.
University of Texas at Austin
Austin, TX USA
rebecca.adaimi@utexas.edu

Ka Tai Ho
Dept. of Electrical and Computer Eng.
University of Texas at Austin
Austin, TX USA
katai.ho@utexas.edu

Edison Thomaz
Dept. of Electrical and Computer Eng.
University of Texas at Austin
Austin, TX USA
ethomaz@utexas.edu

***Abstract*—Ecological Momentary Assessment (EMA) is a data collection method that consists of asking individuals to answer questions pertaining to their behavior, feelings, and experiences in everyday life. While EMA provides benefits compared to retrospective self-reports, the frequency of prompts throughout the day can be burdensome. Leveraging advances in speech recognition and the popularity of conversational assistants, we study the usability of an EMA interface specifically aimed at minimizing the interruption burden caused by EMA. The interface delivers prompts verbally and captures responses with a hands-free voice-based interface, effectively eliminating the need for participants to shift their attention away from a primary task to interact with a mobile phone or smartwatch. In a two-week qualitative study with 13 participants, 69.2% of them reported that not much effort was required to answer prompts with the interface, and 76.9% agreed that the interface helped them integrate EMA into the rhythm of daily life.**

***Index Terms*—ecological momentary assessment, voice input interface, data collection, data annotation, experience sampling, speech recognition**

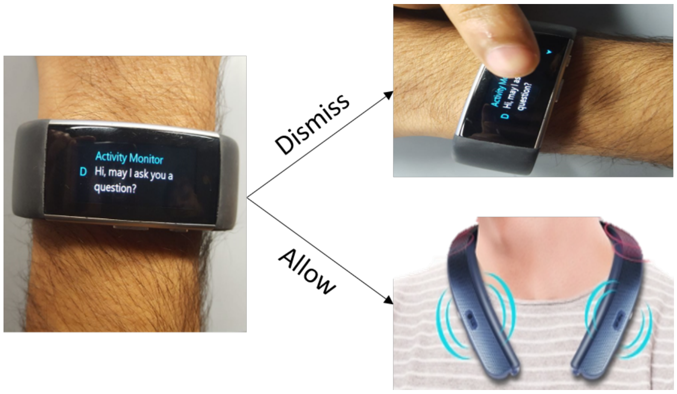


Fig. 1: Our interface requires a wearable speaker for audio input and output, and a wristband which can be optionally used to block incoming EMA prompts.

I. Introduction

Researchers in a wide range of disciplines such as clinical psychology, behavioral science, economics, and ubiquitous computing, are often interested in studying people in natural settings. These studies typically involve querying individuals about their experiences, activities, preferences, feelings and thoughts in daily life. The Ecological Momentary Assessment (EMA) technique is an approach that has been traditionally used to continuously prompt people to respond to questions throughout the day, such as ”how are you feeling right now?” or ”have you interacted with anyone in the last hour?” [1]. This method has proven remarkably useful in studies ranging from cigarette smoking cessation and relapse [2], the relationship between mood and binge eating [3], and adaptation processes related to health and behavioral medicine [4]. Despite the advantages of EMA over self-reports, which are known to have numerous shortcomings, prompting individuals to respond to questions in naturalistic settings can be highly burdensome. A continuous stream of interruptions can lead to bias in the data being collected, or cause participants to ignore prompts altogether [5].

In this paper, we present an EMA interface aimed at minimizing the interruption burden caused by EMA prompting. Our method delivers EMA prompts verbally and captures responses with a hands-free voice-based interface, effectively eliminating the need for participants to shift their attention away from a primary task to interact with a mobile phone or smartwatch. This approach, which aspires to make the process of answering EMA queries in everyday settings *immediate* and *effortless* [6], draws from advances in speech recognition and the emerging popularity of voice-based interfaces (e.g., Amazon Echo, Google Home). These conversational systems allow people to reliably communicate with machines without any direct physical interactions, which is a useful attribute when developing EMA strategies that minimally impact people in the midst of their everyday activities.

The contributions of this work are twofold. Firstly, we demonstrate an implementation of our approach, which consists of a wristband for haptic feedback and control; a wearable speaker for audio input and output; and a workflow for prompt notification and confirmation. While voice-based EMA interfaces exist, ours is the first to offer an end-to-end voice experience, from prompting to response capture. Secondly, we report the results of a usability study of our proposed interface and discuss key challenges and opportunities. In the study, 13 participants used our system over 2 weeks to respond to common EMA questions. It is important to note that our aim with this work is not to advance a new EMA method and test its performance with metrics like compliance, completion and response rates. Instead, we evaluate the suitability of the proposed interface in terms of how usable it is for EMA, and how it is perceived by individuals in this context.

II. Background and Related Work

For many years, researchers have explored numerous ways to lower the interruption burden of EMAs. Today, smartphones have become the de-facto platform for EMAs, but responding to survey questions continues to be an onerous undertaking. Recently, researchers have begun exploring the potential of wearables such as smartwatches for EMA data collection. For example, Intille et al. developed a smartwatch-based EMA system called (μEMA) [7], [8]. The motivation for this work was to optimize prompt response by leveraging a smartwatch and one-touch interactions. Results were quite positive, and later confirmed by Hernandez et al., who compared a smartwatch, a head-mounted device (i.e., Google Glass), and a mobile phone [9]. In the context of stress measurement, the smartwatch prevailed, enabling fast interaction and minimizing burden due to the fact that a smartwatch is more concealable and accessible during the day. An important finding was that accessing the mobile phone every time proved disruptive, especially when the user needs to physically find and reach for the device, such as when it is inside a pocket or purse.

Voice-based approaches have been successfully used for EMA data collection in the past as well. In particular, a number of studies have relied on Interactive Voice Response (IVR) methods [10], [11]. These systems either accept calls from participants or make phone calls at pre-determined times. Once a call is established, an automated system administers surveys verbally; participants respond to questions either verbally as well or using the phone's keypad. While these voice-centric approaches facilitate data collection through a consistent and natural interface, they still require individuals to direct attention to a physical device, i.e., the phone, and away from a primary task.

More similar to our work is the study by Scholl et al., who investigated voice input with Google Glass as a way to label activities [12]. However, it is difficult to generalize its findings, as the evaluation lasted only one 1-hour and did not make use of actual EMA prompts.

III. Implementation

Our approach was implemented with two primary devices: a wearable speaker and a wristband. The wearable speaker, i.e., the Bluetooth-based LG Tone Studio HBS-W120 speaker, is worn around the neck, as shown in Figure 1. The speaker outputs computer-generated voice prompts with EMA questions and captures spoken answers with a built-in microphone. The speech-to-text and text-to-speech conversion was implemented using the Google Cloud Speech API [13]. The wristband, i.e., Microsoft Band 2, can be worn on any wrist and is used to give individuals haptic feedback when the system is ready to capture a voice response following a prompt. Through its touchscreen, the wristband also lets individuals explicitly cancel an incoming prompt, since there are situations when audio output might not be appropriate, e.g. in a meeting, classroom, religious service, or noisy environment. A smartphone was connected to both devices and was used to program and coordinate their interaction:

1) The wristband vibrates when a scheduled voice prompt is imminent, and allows the individual to dismiss it with a simple tap if needed (e.g., the participant is in a meeting and cannot speak). If no action is taken after 10s, the prompt proceeds by asking the scheduled question via the wearable speaker.
2) The wristband vibrates after asking the question to signal when the device is ready to receive the input response. The response is captured through the microphone in the wearable speaker. A 40-second timer is implemented during which an audio response is expected. If it times out and no audio response is captured, the event is logged as a noisy environment or the device was not able to pick up any input audio.
3) When an audio response is received, the system asks for a confirmation of whether the response it captured is accurate, after which the wristband vibrates again to indicate when the system is ready to receive a confirmatory answer with a ”Yes” or ”No”. If the answer is ”No”, the participant is asked to repeat the response. At the end of the prompt and after the user confirms with a ”Yes”, the wristband vibrates to indicate that the response was successfully received.

Figure 2 illustrates the sequence of events. The event workflow and timer durations were refined through an iterative process and informed by a formative study with 3 eligible study participants.

IV. Usability Study

We conducted a usability study to evaluate our proposed EMA interface. In the study, 13 participants (7 males and 6 females, between the ages of 20 and 25) answered EMA questions over a period of 2 weeks. Participants were recruited from the student population of a large public college campus through word of mouth. At the end of the study, participants completed a survey including Likert-scale type questions and provided feedback about their experience in free-form text.

The EMA questions in the study were randomly selected from a set of common questions spanning a range of topics, from dietary habits to mood. Example questions included *”who are you with right now?”*, and *”are you happy right now?”*. Prompts were scheduled at 1-hour intervals.

Typically, smartphone-based EMA questions are restricted to multiple choice questions or visual analog scale questions. Since our approach employs a voice-based response capture method, participants could provide as much detail as desired. For instance, the question *”How stressed are you right now?”* could be answered directly, e.g., yes or not, or with a description of the *cause* of stress, such as a difficult day at work or an upcoming exam.

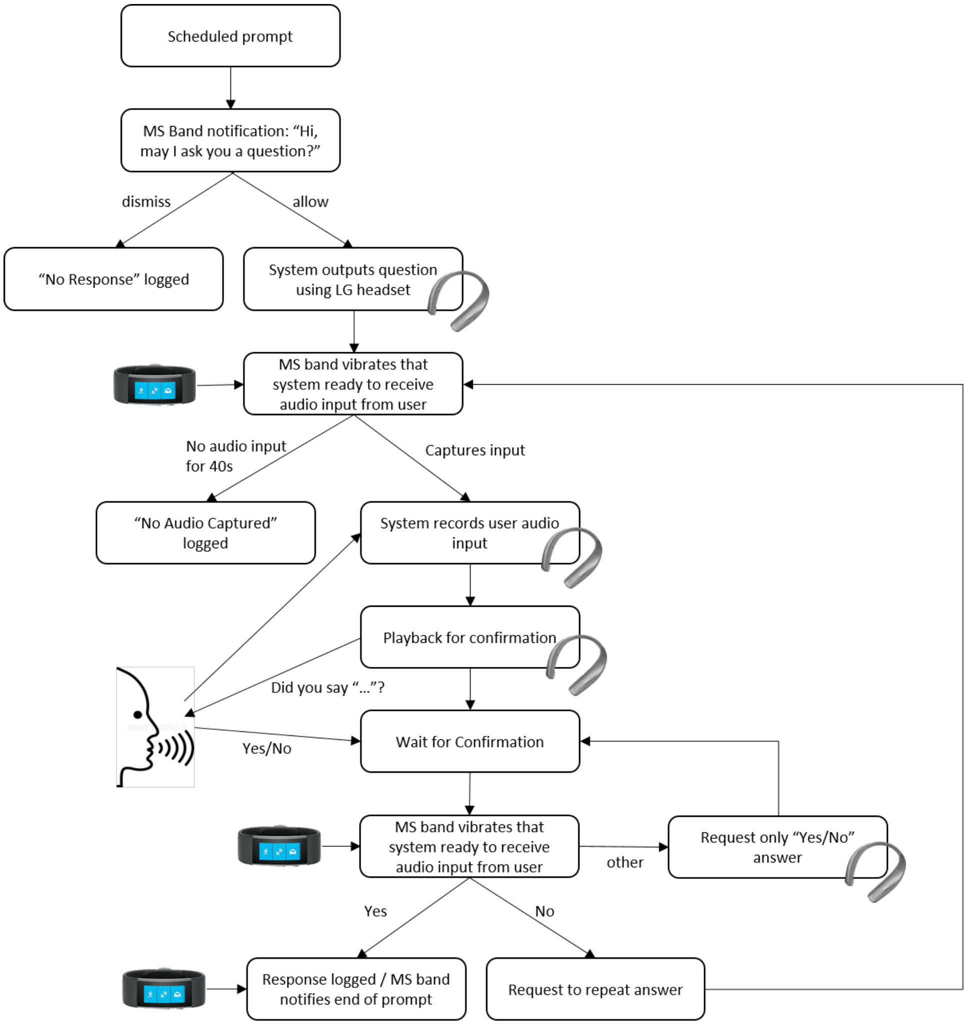

Scheduled prompt
MS Band notification: “Hi, may I ask you a question?”
dismiss
allow
“No Response” logged
System outputs question using LG headset
MS band vibrates that system ready to receive audio input from user
No audio input for 40s
Captures input
“No Audio Captured” logged
System records user audio input
Playback for confirmation
Did you say “...”?
Yes/No
Wait for Confirmation
MS band vibrates that system ready to receive audio input from user
other
Request only “Yes/No” answer
Yes
No
Response logged / MS band notifies end of prompt
Request to repeat answer

Fig. 2: Sequence of events including prompt control, notification and verification

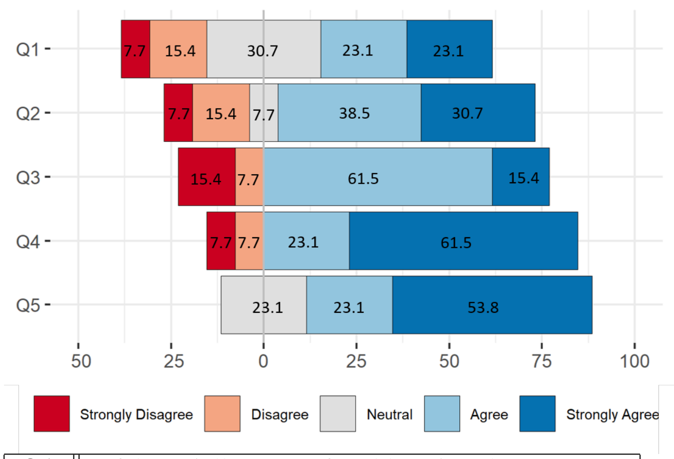
V. Results and Discussion

Our study results, shown in the Likert Scale plot in Figure 3, surfaced both positive and negative aspects of our interface. As we had hoped, participants reported that the EMA prompts were not too distracting (e.g., *”It does not disrupt my daily activities to collect information from me”*, *”Very non-intrusive”*) and the interface was adequate (e.g., *”It was fast and easy most of the time”*). Moreover, the hands-free design provided its intended benefit (e.g., *”I found it easy to respond and was also able to provide more context, e.g. ”Yes, I’m happy because...”, which I otherwise wouldn’t if I were required to type into my phone”*). On the other hand, participants complained about the reliance of the interface on two devices (e.g., *”having to use two wearable devices - would have preferred to only use one”*) and reported dislike for the wearable speaker (e.g., *”I didn’t like having to wear the big headphone thing around my neck the whole time and having that as my only thing to be able to use to respond”*). In the sections below, we discuss the results in Figure 3 and our findings in more detail.

A. Perception of Effort and Burden

A key point of interest is how participants perceive the effort and burden of the interface. 69.2% (Q2: 30.7% (Strongly Agree) + 38.5% (Agree)) of participants reported that not much effort was required. However, some participants had to put some effort to correct captured responses. This is a common challenge in conversational assistants today, e.g., understanding different dialects and accents mainly due to the accuracy of the speech recognition algorithm.

We also hypothesized that a voice-based, hands-free EMA interface would allow participants to answer EMA queries even while performing other activities. 76.9% (Q3: 15.4% (Strongly Agree) + 61.5% (Agree)) of participants indicated this was indeed possible and appreciated this capability of the system (e.g., *”I liked that it was hands-free - I could respond while I was driving or while I was studying and didn’t really have to stop what I was doing to use it”*).



Q1	I felt socially uncomfortable using the system.
Q2	Responding to prompt did not require much effort.
Q3	I was able to perform my everyday activities while using the system.
Q4	The wristband was useful for blocking prompts.
Q5	I found the wristband vibrations to be useful in providing guidance for voice input.

Fig. 3: Likert Scale Plot of Online Survey

B. *Prompt Notification, Control and Verification*

A wristband was used as a control interface to dismiss prompts, deliver notifications and guide voice input. 84.6% (Q4: 61.5% (Strongly Agree) + 23.1% (Agree)) of participants agreed that the wristband was useful for blocking the prompts (e.g., *"I liked having the ability to block prompts"*). Moreover, 76.9% (Q5: 53.8% (Strongly Agree) + 23.1% (Agree)) found the device useful in providing guidance for voice input via the haptic feedback (e.g., *"The vibrations were a good indication"*).

Additionally, our interface included an acknowledgement step that asks the participant to confirm whether the system captured their response correctly. This crucial step was added as a solution to the problem that most voice-based systems suffer from, clearly capturing a user's response. Unfortunately, we observed that this also had a negative impact on the quality of data. Since having to repeat the answer again adds burden to the user, some participants admitted to confirming the response even when it was wrong just to avoid having to repeat their response again. For instance, for the question *"How stressed are you right now?"*, some invalid responses included *"not that old"*, which at that moment, the participant reported trying to say *"not at all"*.

To improve intelligibility in the audio exchange, some participants suggested using the wristband's display to complement voice input and output (e.g., *"Felt fine and worked well, I think the only update might be displaying the question asked on the screen in case you missed what the speaker said"*).

C. *Social Acceptability and Privacy*

Almost half of the study participants, 46.2% (Q1: 23.1% (Strongly Agree) + 23.1% (Agree)), expressed feeling socially uncomfortable while using the system in public, and one participant admitted to having only used the system at home or when alone. Upon further analysis, we identified three key factors as the cause for this discomfort. Firstly, due to social norms, communicating to a conversational assistant in public could be disturbing to other individuals nearby or not be allowed at all (e.g., at a library or movie theater).

Second, some individuals reported feeling *awkward* communicating with a machine in public, or disliked wearing the wearable speaker and wristband (e.g., *"It attracted weird attention from my social circle"*, *"I didn't like having to wear the big headphones, I'm just a little self conscious about that kind of stuff"*, *"I felt awkward talking to the prompter when I was in a quiet room. It was even more awkward when the prompter incorrectly heard my response"*).

While not explicitly mentioned by any participants, the need for privacy is imperative when providing very personal information in EMAs verbally. Our interface allows individuals to block incoming EMA prompts at inopportune times but in some cases this measure might not be enough. However, we see opportunities for addressing this challenge in the future, such as with new types of speech input interfaces, e.g., silent speech [14].

VI. Conclusion

In this paper, we present an EMA interface aimed at improving the experience of responding to EMA prompts in naturalistic settings. In a two-week usability study, our hands-free voice-based interface proved promising as the foundation for novel EMA methods that integrate seamlessly with daily activities. As per our results, a large percentage of our participants agreed that not much effort was required to respond to prompts. Although we also received negative feedback about the interface and experience, we foresee many paths for addressing the reported shortcomings of our approach, including improvements to our prompt notification and verification workflow, and the utilization of different hardware platforms for the wearable speaker and wristband. Moreover, a longer and more comprehensive user study is needed to further prove the usability of our EMA interface due to our limited assessment setting in terms of number of participants and study duration. Therefore, we intend to address these issues and conduct further studies in future work.

References

[1] J. Cain, Depp, "Ecological momentary assessment in aging research: A critical review," *Journal of Psychiatric Research*, vol. 43, no. 11, pp. 987–996, 2009. [Online]. Available: http://doi.org/10.1016/j.jpsychires.2009.01.014

[2] S. Kirchner and Wileyto, "Relapse dynamics during smoking cessation: Recurrent abstinence violation effects and lapse-relapse progression," vol. 121, pp. 187–197, 02 2012.

[3] K. E. Wegner, J. M. Smyth, R. D. Crosby, D. Wittrock, S. A. Wonderlich, and J. E. Mitchell, “An evaluation of the relationship between mood and binge eating in the natural environment using ecological momentary assessment,” *International Journal of Eating Disorders*, vol. 32, no. 3, pp. 352–361, 11 2002. [Online]. Available: http:https://doi.org/10.1002/eat.10086

[4] A. Stone and S. Shiffman, “Ecological momentary assessment (ema) in behavioral medicine,” vol. 16, pp. 199–202, 01 1994.

[5] S. Shiffman, A. Stone, and M. Hufford, “Ecolocial momentary assessment,” vol. 4, pp. 1–32, 02 2008.

[6] T. Starner, “Project glass: An extension of the self,” *IEEE Pervasive Computing*, vol. 12, no. 2, pp. 14–16, 2013.

[7] S. Intille, C. Haynes, D. Maniar, A. Ponnada, and J. Manjourides, “μema: Microinteraction-based ecological momentary assessment (ema) using a smartwatch,” in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp ’16. New York, NY, USA: ACM, 2016, pp. 1124–1128. [Online]. Available: http://doi.acm.org/10.1145/2971648.2971717

[8] A. Ponnada, C. Haynes, D. Maniar, J. Manjourides, and S. Intille, “Microinteraction ecological momentary assessment response rates: Effect of microinteractions or the smartwatch?” *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 3, pp. 92:1–92:16, Sep. 2017. [Online]. Available: http://doi.acm.org/10.1145/3130957

[9] J. Hernandez, D. McDuff, C. Infante, P. Maes, K. Quigley, and R. Picard, “Wearable esm: Differences in the experience sampling method across wearable devices,” in *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, ser. MobileHCI ’16. New York, NY, USA: ACM, 2016, pp. 195–205. [Online]. Available: http://doi.acm.org.ezproxy.lib.utexas.edu/10.1145/2935334.2935340

[10] R. L. Collins, T. B. Kashdan, and G. Gollnisch, “The feasibility of using cellular phones to collect ecological momentary assessment data: Application to alcohol consumption.” *Experimental and clinical psychopharmacology*, vol. 11, no. 1, p. 73, 2003.

[11] D. T. Duncan, W. C. Goedel, J. H. Williams, and B. Elbel, “Acceptability of smartphone text-and voice-based ecological momentary assessment (ema) methods among low income housing residents in new york city,” *BMC research notes*, vol. 10, no. 1, p. 517, 2017.

[12] P. M. Scholl, M. Borazio, M. Jänsch, and K. Van Laerhoven, “Diary-like long-term activity recognition: Touch or voice interaction?” in *2014 11th International Conference on Wearable and Implantable Body Sensor Networks Workshops*. IEEE, 2014, pp. 42–45.

[13] “Cloud speech-to-text - speech recognition — cloud speech-to-text api — google cloud,” Oct 2017. [Online]. Available: https://cloud.google.com/speech-to-text/

[14] A. Kapur, S. Kapur, and P. Maes, “Alterego: A personalized wearable silent speech interface,” in *23rd International Conference on Intelligent User Interfaces*, ser. IUI ’18. New York, NY, USA: ACM, 2018, pp. 43–53. [Online]. Available: http://doi.acm.org/10.1145/3172944.3172977