Capturing Voices at Risk : Designing Semi-Structured Speech Tasks for Voice Data Collection in Depression and Anxiety Risk Groups

Maryam Hemati[1], Woojin Seung[1], Sohyun Jun[1], Jee Hang Lee[2,3*], Jinwoo Kim[4]

[1]Department of Cognitive Science, Master candidate, Yonsei University, Seoul, Korea

[2]Department of Human-Centered AI, Assistant Professor, Sangmyung University, Seoul, Korea

[3]Institute for Advanced Intelligence Study, Daejeon, Korea

[4]Department of Cognitive Science, Professor, Yonsei University, Seoul, Korea

Abstract

Background The human voice carries rich multimodal information, disclosing cognitive and emotional states through linguistic content and acoustic properties. Voice data has emerged as a promising digital biomarker for detecting anxiety and depression in at-risk populations. However, collecting high-quality voice data that effectively captures vocal manifestations of psychological distress remains a significant methodological challenge.

Methods We developed “MindCheck,” a conversation-based semi-structured speech task designed specifically for at-risk depression and anxiety populations. Our two-phase approach first collected user requirements through interviews to establish user experience (UX) desiderata for voice data acquisition, then designed and validated a human subject task to capture informationally rich voice data from individuals at risk.

Results Iterative task refinement improved user engagement and reduced task completion difficulties. Participants emphasized the importance of engaging interactions and clear guidance, which resulted in more adjustable interface configurations. Higher engagement levels and more natural interaction patterns suggest improved circumstances for capturing authentic voice biomarkers.

Conclusions This study highlights the critical role of user-centered design in developing effective voice data collection protocols for anxiety and depression risk assessment. The iterative task refinement approach demonstrated that prioritizing user engagement creates better circumstances for authentic voice biomarker capture, enhancing the potential of conversation-based tools for early detection in vulnerable populations.

Keywords Semi-structured Speech Task, Voice Biomarker, Task Design, mHealth, Anxiety and Depression

This study was supported by the 2024 Seoul R&BD Program (Grant No. CY240047) and by the Ministry of Trade, Industry and Energy (MOTIE), Korea, under the Bio-Industrial Technology Development Program (Project No. RS-2024-00431485), supervised by the Korea Evaluation Institute of Industrial Technology (KEIT).

*Corresponding author: : Jee Hang Lee (jeehang@smu.ac.kr)

Citation: Hemati, M., Seung, W., Jun, S., Lee, J., & Kim, J. (2025). Capturing Voices at Risk : Designing Semi-Structured Speech Tasks for Voice Data Collection in Depression and Anxiety Risk Groups. *Archives of Design Research, 38*(4), 51-67.

http://dx.doi.org/10.15187/adr.2025.11.38.4.51

Received : Feb. 24. 2025 ; **Reviewed :** Sep. 17. 2025 ; **Accepted :** Sep. 17. 2025
pISSN 1226-8046 **eISSN** 2288-2987

1. Introduction

Mental health disorders are among the leading causes of disability worldwide, yet traditional methods of assessment – such as self-reported surveys – suffer from inherent limitations, including subjectivity, underreporting, and lack of ecological validity [31]. In recent years, the use of digital biomarkers has emerged as a promising alternative for monitoring emotional and psychiatric states. Among these, voice and heart rate variability (HRV) have gained traction due to their passive, non-invasive nature and compatibility with mobile health (mHealth) systems [28,44].

Voice biomarkers, in particular, offer rich, multidimensional signals tied to cognitive, emotional, and physiological states. Research has shown that vocal characteristics such as pitch, speaking rate, spectral features, and vocal jitter can serve as indicators of depression, anxiety, and suicidal ideation [8,16,17,27,48]. For example, reduced pitch variability and slower speech rates are often observed in depression, while anxiety is associated with increased pitch and vocal disfluencies [2,18,19,27,33,55,56,59]. However, the effective use of these digital biomarkers requires high-quality, contextually rich datasets, which are often lacking in clinical and real-world settings [4, 23].

Most existing datasets rely on either scripted readings or unstructured conversations, thus limiting their generalizability and clinical utility [23, 36]. To address these limitations, we propose a user-centered framework for designing semi-structured speech tasks that encourage diverse, emotionally expressive, and diagnostically informative voice recordings.

This work is part of developing *MindCheck*, a mobile-based digital healthcare application that integrates multiple data sources including structured questionnaires, voice input, and heart rate variability via remote photoplethysmography (rPPG). In this paper, we specifically focus on the additional task design for voice data collection.

Using a mixed-methods approach involving literature review, expert input, and iterative testing, we developed a series of speech tasks aimed at eliciting meaningful vocal responses across emotional and cognitive dimensions. Task-based designs enhance user engagement and data quality in speech-based assessment systems [47].

We evaluated these tasks through two stages: a pre-study involving clinical question mapping and expert interviews, and a main study testing user experience and usability with a mobile interface. Our ultimate goal is to enhance the reliability of speech-based digital biomarkers in mental health research by advancing task design. This approach establishes the foundation for future diagnostic modeling by ensuring quality and richness of collected data through enhanced user engagement.

2. Method

To develop novel semi-structured speech tasks, we conducted research in two stages: pre- and main study (Figure 1). This allowed us to iteratively design and develop a voice-based conversational task for collecting informationally rich data required for mental health assessment.

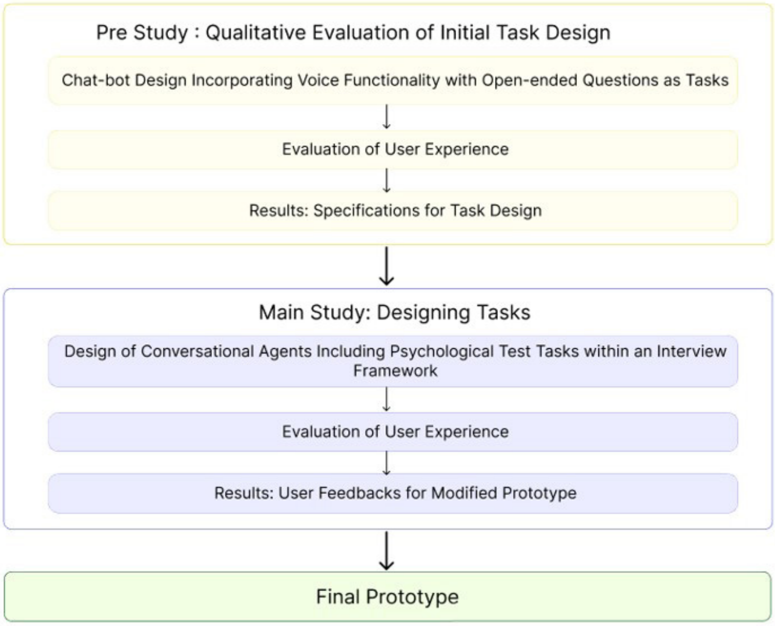


Figure 1 Diagram of Study Structure.

After the initial task design, we conducted a pre-study using voice functionality and open-ended questions to qualitatively evaluate the design. This approach allowed us to incorporate findings from relevant literature while also gathering in-depth user feedback. Evaluating the user experience provided us with valuable insights, which we then used to develop task enhancement specifications that aligned with both academic research and practical usability needs.

Building on these findings, the main study further developed the system by creating a conversation-based task that incorporated psychological assessments into an interview framework. This stage focused on combining insights from existing literature with specific user feedback to ensure the system's effectiveness, usability, and relevance. We systematically collected user feedback on the refined design, which enabled us to make iterative improvements.

Pre-Study: Qualitative Evaluation of Initial Task Design

Participants

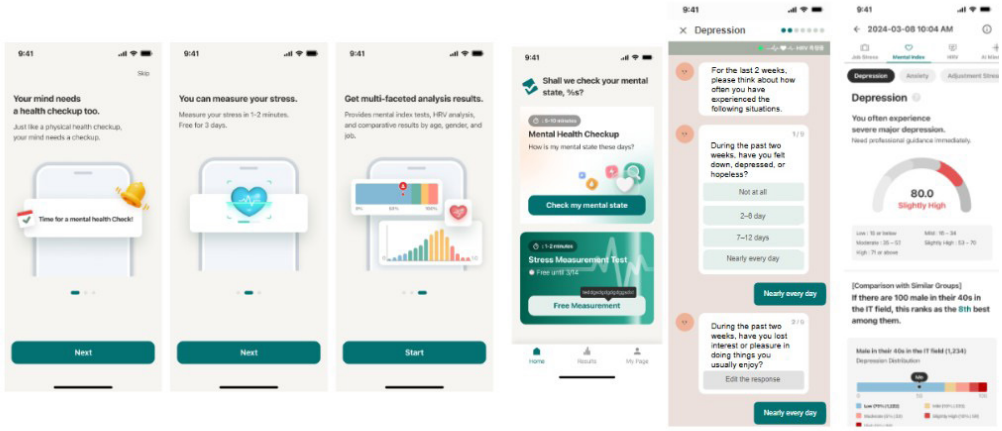
Nine participants (five males; mean age: 36.3 years, SD = 5.2) assessed the preliminary *MindCheck* prototype with its innovative voice feature (for more details on participants, see Supplementary Table 1). Participants were recruited through online invitations on social media platforms. All participants performed full *MindCheck* tasks, which included seven clinical questionnaires, HRV measurement, and voice-recording tasks conducted at multiple stages.

MindCheck: A mobile-based digital healthcare application

MindCheck is a mobile application for mental health assessment that digitally delivers standardized clinical questionnaires and simultaneously measures heart rate variability (HRV) using a remote photoplethysmography (rPPG) technology [35]. Aforementioned, the application provides an integrated screening process, combining psychological self-report instruments with physiological monitoring. The digital assessment component includes: (1) Patient Health Questionnaire-9 (PHQ-9) for depressive symptoms [30] , (2) Generalized Anxiety Disorder-7 (GAD-7) for anxiety [53] , (3) Adjustment Disorder New Module-4 (ADNM-4) for adjustment disorder [20] , (4) Primary Care PTSD Screen for DSM-5 (K-PC-PTSD-5) for post-traumatic stress disorder [45] , (5) Insomnia Severity Index-Korean (ISI-K) for insomnia [13], (6) Columbia-Suicide Severity Rating Scale (C-SSRS) for suicidality [49], and (7) the Korean Occupational Stress Scale (KOSS) for occupational stress [12].

These questionnaires are delivered through a conversation-based user interface (UI), which guides users through each question in a dialogue format. The chatbot was designed based on Shevat's five-step framework [51]. During questionnaire interaction, the application activates the front camera of the mobile device to detect subtle facial blood flow changes, enabling contactless HRV estimation via rPPG. This simultaneous recording provides complementary physiological data that may reflect emotional arousal or stress responses during the psychological assessment [39].

Beyond these functionalities, we added a naturalistic conversational interface, called PAPA. This interface was designed to express adaptive personality traits, informed by the Five Factor Model [52], and tailored to user demographics (e.g., age, gender) to improve rapport and authenticity in responses. Additionally, we included a "camera mirror" feature, during the interactions between PAPA and users. This design element draws from prior work demonstrating that visual feedback from camera-based physiological systems enhances user reflection and comfort, as seen in mirror-mediated affective interfaces that connect outer appearance with inner state visualizations [26].



(A) Onboarding pages (B) Home (C) Chatbot page (D) Result page

Figure 2 Example pages of MindCheck.

Interview Process and Question Design

Given the *MindCheck* application described above, participants were required to complete the digital assessment. Subsequently, a structured qualitative interview was conducted to assess their perspectives on usability, emotional engagement, and voice interaction design. These interviews were guided by feedback from the initial prototype with added tasks, and explored how participants experienced the onboarding process, the voice task, and the results interface.

The interview began with questions focused on the onboarding experience. We examined whether the introductory explanation was sufficiently clear, especially regarding the distinction between the survey and HRV/voice biomarker components. Participants were asked whether presenting these assessments as two separate but complementary modules affected their perception of burden or task length. Additional prompts explored users' impressions of the explanation content and tone (UX writing), and whether the introductory mention of emotion recognition through voice felt engaging or persuasive. Participants who found it less appealing were asked to suggest alternative ways to frame the purpose or value of voice analysis.

Next, the interview addressed the voice recording phase guided by PAPA. Questions explored whether the microphone check process caused discomfort, and how the system could better encourage natural speech. Participants were also asked whether combining a conversation-based approach with the voice button in one interface created usability friction. To assess motivational alignment, they were prompted to describe what outcomes they expected or valued from voice analysis, and whether moving the voice task earlier in the session (e.g., as part of the opening) would be perceived as more or less cognitively demanding.

Three key open-ended questions were placed at distinct stages of the interaction to serve as structured voice tasks. First, prior to the start of the survey tasks, participants were asked, "How was your day today?" – a neutral warm-up question often used in qualitative psychological interviews to facilitate conversational flow and lower emotional inhibition [9].

The second question – “When you feel anxious or stressed, what are the activities you do to cope with it?” – was asked after participants completed PHQ-9, GAD-7, and KOSS. This prompt was informed by emotion regulation and coping literature [11], providing insight into real-life behavioral responses to stress and how these may relate to participants’ engagement with the app content.

At the end of the session, participants were asked: “Do you have any feedback or suggestions regarding the app experience?” This final reflection was aimed at eliciting feedback on both content and interaction design, in line with human-computer interaction and digital health usability research [7]. Participants were also invited to share any additional thoughts about the interview or overall experience.

After reviewing the results screen, participants were asked about their experience during the wait time specifically – whether it felt passive and whether certain interface elements could reduce the sense of waiting. The emotional space graph was also evaluated for clarity and intuitiveness.

In the final segment, participants were asked two key questions: first, whether they would be willing to complete the voice task independently of the full mental health screening; and second, what kind of results presentation or UX writing would make voice-based feedback feel valuable. Finally, they provided feedback on desired UI improvements for the onboarding, recording, and results stages, and were encouraged to share thoughts on how the emotional space graph might offer additional value or be improved.

The full list of interview questions and their corresponding phases are presented in Supplementary Table 2.

Main Study: Designing Tasks

Procedure

Based on the findings from the pre-study, the main study was designed to develop and evaluate a set of speech tasks capable of eliciting high-quality, informationally-rich voice data, while maintaining a strong emphasis on user comfort and engagement. The task design process involved several key phases.

First, task requirements were derived directly from the pre-study interviews. Next, we conducted an extensive review of the literature on clinical psychology, digital mental health, and voice-based diagnostic tools to identify task types with strong theoretical and empirical grounding [1, 10,19,24,32,37,60]. Task selection was guided by two primary criteria: (i) psychological relevance [1], meaning that participant responses would plausibly reflect underlying emotional or mental health states; and (ii) structural diversity [19,37], ensuring inclusion of structured, semi-structured, and free-form tasks to enable downstream analysis of both emotion regulation patterns and disorder-specific response features.

Following the task selection, an initial set of prototypes – both in task content and user interface – was developed. These prototypes were implemented into a working version of the application and evaluated through pilot interviews with users. Feedback from these sessions was used to iteratively refine both the interaction design and the voice task content.

The final prototype integrated principles from conversational user interface research, mental health technology design, and digital therapeutics. It specifically targeted improvements in the quality of voice data – measured by content richness and speech length – while enhancing the usability and engagement of the overall experience. The refined version is positioned for evaluation in future research phases, where its utility for digital mental health assessment will be more comprehensively tested.

Ultimately, the final set of speech tasks was selected and adapted based on evidence from established mental health diagnostics and prior research in voice-based screening. Each task was carefully constructed to evoke emotionally and cognitively meaningful speech, ensuring compatibility with both acoustic and linguistic analysis methods. The design emphasized the collection of high-quality voice data while safeguarding participant comfort. Linguistically optimized questions were employed to elicit speech features relevant to key mental health indicators, drawing directly from validated diagnostic frameworks and research protocols. This balance between diagnostic relevance and user-centered delivery was central to preserving engagement and minimizing cognitive load throughout the assessment process.

Evaluation

The evaluation protocol engaged 10 participants (3 males; mean age: 28.3 years, SD = 4.1) recruited through social media channels. Participants received an initial MindCheck briefing prior to system engagement. The protocol excluded supplementary voice function guidance to simulate naturalistic usage conditions, contrasting with the pre-study methodology.

The assessment permitted unrestricted system interaction, enabling evaluation of response quality and audio duration metrics. Post-task semi-structured interviews examined functional efficacy and usability factors. Researchers analyzed interview data through paragraph coding and categorical classification to identify user experience patterns.

3. Result

3. 1. Implications from the Pre-Study: Requirements for the Main Task

The pre-study findings resulted in critical insights for enhancing the task structure and voice biomarker collection strategies within the *MindCheck* application. Notably, sequencing the voice prompt after participants completed the clinical surveys proved to be highly effective. This timing allowed participants to first become familiar with the app's purpose and emotional context, resulting in more thoughtful and contextually grounded responses, as participant P03 noted: *"If I had to speak right at the beginning, I wouldn't know what to say. After the surveys, I had a clearer picture of what this app is about."*

When voice data were collected in isolation, without preceding contextual engagement, responses tended to be brief and superficial. In contrast, a dialogue-based structure elicited richer and more expressive speech samples. Seven out of nine participants expressed a clear preference for a conversational interface, noting that it enhanced the naturalness and emotional comfort of their responses. For example, P06 commented, *"It felt more like talking to someone, not just answering a form."*

Preferences regarding input modality varied across users. Four participants preferred text-only presentations, citing privacy or clarity concerns: "*Sometimes I'd rather read quietly than speak out loud, especially in public* (P04)." Three preferred voice-only interactions, stating it felt more natural or effortless: "*Just talking without reading feels easier for me when I'm tired* (P02)." Two participants advocated for having both options available depending on the user's context or state: "*Sometimes I want to read, sometimes I want to hear it—it depends on my mood or surroundings* (P09)." These responses suggest the importance of offering dual-modality interfaces for personalized and context-sensitive user experiences.

All participants unanimously highlighted the importance of interactive visual elements, such as progress bars, animated transitions, or mood-based visuals, as essential to sustaining attention and reducing perceived repetitiveness. P08 remarked on this: *"Even simple animations or visual changes made it feel less like a chore and more like a guided experience."*

Moreover, participants emphasized the value of receiving tangible feedback based on their voice input. Several users expressed that understanding how their data was being interpreted increased their sense of trust and motivation to engage. For instance, P01 stated: *"It would be helpful to know what my voice says about my mood – even just a simple summary."* This interest in actionable feedback reflects a strong preference for reciprocity in digital health tools and suggests that perceived transparency can elevate user satisfaction.

The findings also highlighted the importance of high-quality voice data in developing precise and reliable digital biomarkers. Participants responded more deeply when questions were designed to evoke emotional expression and when they were situated within an intuitive and thoughtfully sequenced interaction flow. The complexity and framing of the voice tasks were shown to directly influence the depth and nuance of the collected speech data.

Taken together, these findings informed several refinements to the *MindCheck* prototype. Design improvements for the next development phase included: (i) enhanced dialogue scaffolding through an interactive digital guide; (ii) optimized task sequencing to promote natural voice elicitation; (iii) integration of visual cues and animations for improved engagement; and (iv) interface flexibility to accommodate varied user preferences in communication mode. These elements collectively contributed to the development of a robust, user-centered system for voice-based mental health assessment. Example screenshots of the initial prototype are presented in Figure 2.

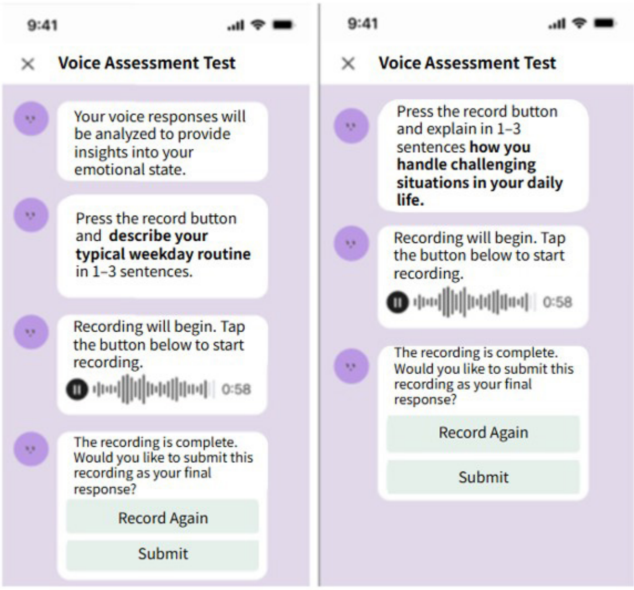


Figure 3 Example of Initial Prototype Used in The Pre-Study.

3. 2. The Outcome of the Main Study

As described, requirements derived from the pre-study included the need for emotionally relevant prompts, support for multiple modalities (e.g., text and voice), structured task sequencing to reduce cognitive burden, and an engaging interface to sustain user attention throughout the interaction. Accordingly, we performed the refinement of the tasks.

Task Design Results

Task 1- Manchester Color Wheel

We incorporated findings from research on the "Manchester Color Wheel," a visual assessment instrument that examines emotional expression through colour selection in patients with depression and anxiety versus healthy controls. This research demonstrated distinct colour preferences for emotional expression between individuals with mental health conditions and those without such diagnoses [10]. Figure 3 illustrates the color wheel used in this assessment.

The protocol requires participants to select colours that correspond to their emotional state and articulate their rationale. This structured approach facilitates colour-choice clustering and justification analysis, which yields insights into psychological states. The methodology is effective through its integration of visual-verbal responses, enabling assessment of mood expression and emotional intensity via vocal data analysis.

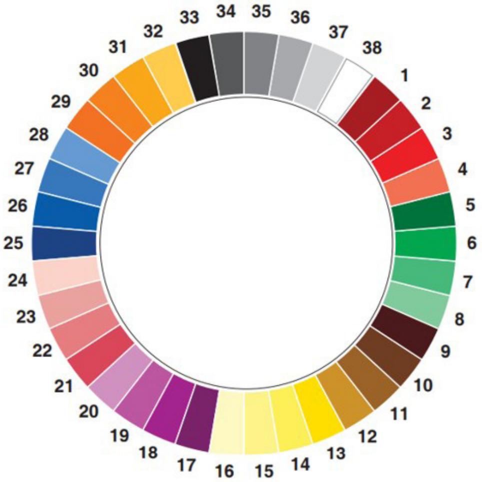


Figure 4 The 'Color Wheel'.

Task 2: WHOQOL-BREF Quality of Life Assessment

Task 2 investigates depression's influence on WHOQOL-BREF domains, examining physical, psychological, and social quality-of-life indicators [60]. The protocol prompts participant evaluations across diverse well-being dimensions, including general health satisfaction, physical comfort, and sleep patterns. This assessment approach tracks response variations, particularly quality-of-life changes following depression intervention, to evaluate mental health trajectories and symptom severity. The integration of Likert scales with qualitative justifications produces structured data suitable for systematic analysis, enhancing mental health insights. We employ empirically validated tasks for targeted data extraction. This approach optimizes mental health diagnostic processes through comprehensive linguistic and acoustic analyses of vocal responses.

Task 3: Twenty Statements Test (TST)

The task evaluates self-concept through participants' responses to the "I am..." prompt, with a twenty-statement limit [32]. Research demonstrates that individuals with depression use more negative self-descriptors, while mentally healthy participants use positive or neutral terminology [24]. This protocol enables systematic response aggregation for examining positive-negative term frequencies. Comparative analysis between healthy individuals and those with depressive symptoms provides insights into emotional states and self-perception. The methodology is particularly effective for identifying depression-linked shifts in self-concept, highlighting its role in mental health assessment. This research framework incorporates empirically validated protocols for targeted data acquisition, optimizing mental health evaluation through comprehensive linguistic and acoustic analyses of vocal responses.

Prototyping

The refined prototype integrated "PAPA," an interactive conversational framework, with enhanced interface elements that addressed pre-study feedback. Key improvements included an improved question sequence, refined voice interactions, and intuitive task designs to enhance user engagement. Prior research demonstrates that well-structured conversational frameworks with adaptive and emotionally responsive voice interactions significantly increase user engagement, perceived empathy, and data richness in mental health applications [6,34]. Moreover, intuitive task flow and minimized cognitive load are critical for maintaining usability and adherence in the digital therapeutic system [3,58]. Figure 4 illustrates the system architecture with multiple functionalities for data acquisition and analysis.

The protocol began with a consent interface ensuring ethical compliance and securing participant agreement. A comprehensive voice function guide provided explicit operational instructions for the voice-based system. The prototype featured three core assessments: (i) the Manchester Colour Wheel for emotional state evaluation, (ii) the WHOQOL-BREF for quality-of-life assessment, and (iii) the Twenty Statements Test for self-perception analysis, as described above. The system concluded with a user experience survey and session summary to collect usability feedback. These integrated elements established a robust framework for voice-based mental health evaluation and insight generation.

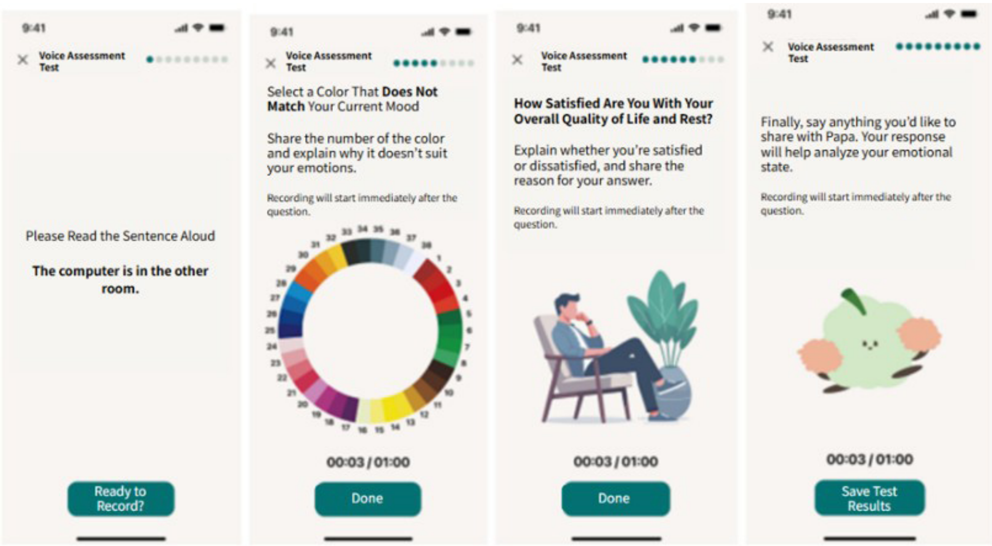


Figure 5 Representative pages of the modified prototype design incorporating the requirements derived from the pre-study.

Qualitative Results on the Newly Designed Tasks

Interview findings demonstrated the voice-enabled system's efficacy and appeal.

Attractiveness Participants found the voice function "fresh and engaging" compared to text-based surveys. P5 explained: "Psychological or stress tests are usually checklist-based self-reports, but speaking through voice feels new and refreshing." P1 highlighted accessibility benefits: "It can be used even by people with visual impairments." Overall, participants emphasized that the voice feature significantly improved both engagement and accessibility.

Agent Persona Participants felt PAPA's voice was comforting and fitting for its cute, approachable character. P2 noted: *"It fit and was easy to answer."* Several users suggested a more professional or sympathetic voice to improve the system's attractiveness. For example, P4 commented: *"It could be warmer and more approachable, but felt less professional."* Similarly, P7 suggested: *"A soft male voice or a low-toned female voice might work better."*

Challenges During the task, participants experienced difficulty with the voice interface speed, with some finding it either too fast or too slow. P1 noted: *"The voice felt a bit too fast,"* while P8 commented: *"It was frustrating to wait for the voice to finish."* Rapid changes in question formats, such as switching from multiple-choice to voice input, were also cited as a challenge. P3 observed: *"Switching formats mid-way was inconvenient."*

Voice input was generally perceived as more time-consuming than text and required participants to adapt to the process, which some found unfamiliar or demanding. P8 explained: *"Listening and speaking takes longer than just typing"* while P2 added: *"The questions made me think deeply, which was unfamiliar."*

Perceptions of Voice Analysis Results Finally, the findings revealed that while participants valued the system's visual output clarity, they sought more comprehensive analytical explanations to enhance comprehension and establish trust in the results. P1 noted: *"The visuals made it easy to understand at a glance."* but P6 suggested: *"It would be better to include reasoning behind the results."* This interview feedback helped refine the prototype by simplifying the interface, improving voice functionality, and enhancing task flow for a more seamless user experience.

Final Prototype

The main interview findings prompted significant prototype refinements. The interactive digital guide *PAPA* adopted a mature female voice profile, replacing the adolescent tone to establish professional rapport.

System usability improved through voice speed optimization to mirror natural conversational cadence. Task protocols underwent streamlining for enhanced clarity. The TST [24,32], initially positioned as the third assessment, proved problematic due to time constraints and complexity, despite the provided prompts. Audio analysis revealed no discernible patterns between varying mental health profiles in MindCheck results, leading to TST removal from the final iteration.

The Manchester Colour Wheel, which garnered the highest participant engagement, received enhancement through the addition of mood-contrast colour selection, thereby improving clustering precision. The protocol incorporated an initial neutral-content baseline task, collecting emotionally neutral vocal data for comparative analysis, drawing from established research methodologies [29].

Figure 5 shows four representative phases of the final prototype across distinct tasks, highlighting the refined voice interface, optimized question sequencing, and enhanced assessment protocols.

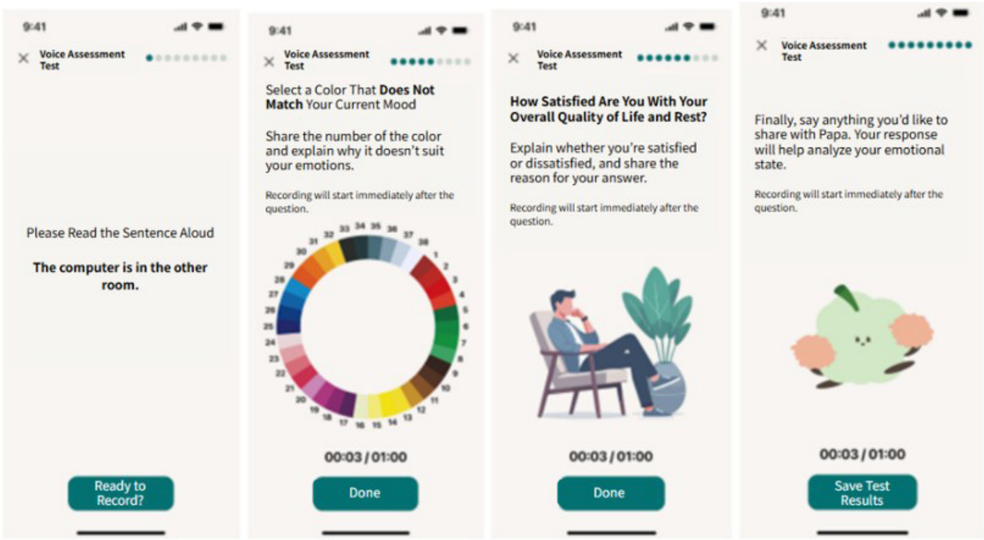


Figure 5 A snapshot of the modified prototype.

Design Rationale for UX and Interaction Structure

The final interface and task flow were grounded in well-established principles of human-computer interaction (HCI) and digital health UX design, with specific emphasis on reducing cognitive load, promoting emotional readiness, and optimizing voice data capture. The task sequence followed a progressive disclosure strategy [43], presenting simpler, emotionally neutral tasks first to minimize early performance anxiety and establish vocal baselines. This approach mirrors emotional acclimatization methods used in mental health technologies to improve task compliance and response quality [40].

The voice interface was deliberately designed using a neutral female tone at a conversational pace, informed by evidence that female voices are perceived as more trustworthy and soothing in healthcare settings [42], and that slower, emotionally congruent speech delivery enhances user trust and self-disclosure [5]. Visual interface elements were kept clean and uncluttered, with dynamic progress indicators and a "camera mirror" mode to reinforce user agency and improve rPPG signal alignment, respectively. This approach aligns with best practices in personal informatics and passive sensing UI, which emphasize transparent feedback and contextual data presentation to foster user trust and long-term engagement [15,21].

The system also adopted multi-modal presentation of task prompts (audio + text) to support accessibility and comprehension, especially for users with differing cognitive styles or attention limitations. This aligns with dual-channel processing theory and has been shown to improve task adherence in digital therapeutic tools [38]. Finally, real-time interaction logging and user feedback mechanisms were incorporated to support future refinement under the V3 framework (Verification, Validation, and Evaluation) [25], ensuring continuous iteration of both user experience and data quality.

4. Conclusion

The enhanced prototype facilitates Korean voice data acquisition for future machine learning model training. This dataset will be instrumental in optimizing real-time acoustic and linguistic analysis capabilities, initially focusing on prevalent conditions such as depression and anxiety. Our research trajectory envisions a comprehensive monitoring framework that transcends diagnostic boundaries, enabling intuitive mental health tracking while circumventing traditional self-reporting constraints and enhancing accessibility and user engagement.

This research demonstrates the critical importance of effective task design in voice biomarker implementation for mental health evaluation. The refined MindCheck system presents a non-invasive, scalable methodology for identifying depression, anxiety, and suicidal ideation through acoustic parameters and linguistic features. Our two-phase investigation informed the development of an optimized interactive digital guide and task architecture that incorporates validated assessment protocols. By integrating iterative task designs with high-quality data collection, we demonstrate voice biomarkers' potential for advancing digital therapeutics. Future Korean data collection will enhance analytical model precision, with the ultimate goal of facilitating intuitive mental health monitoring that fosters self-awareness while minimizing reliance on conventional assessment instruments.

References

1. Aldeneh, Z., Jaiswal, M., Picheny, M., McInnis, M., & Provost, E.M. (2019). Identifying mood episodes using dialogue features from clinical interviews. *arXiv preprint arXiv:1910.05115*.
2. Amato, F., Rechichi, I., Borzì, L., & Olmo, G. (2022, March). Sleep quality through vocal analysis: a telemedicine application. In *2022 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)* (pp. 706-711). IEEE.
3. Balcombe, L., & De Leo, D. (2022, February). Human-computer interaction in digital mental health. In *Informatics* (Vol. 9, No. 1, p. 14). MDPI.
4. Bensoussan, Y., Elemento, O., & Rameau, A. (2024). Voice as an AI biomarker of health-introducing audiomics. *JAMA Otolaryngology-Head & Neck Surgery, 150*(4), 283-284.
5.Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI), 12*(2), 293-327.
6. Blandford, A., Furniss, D., & Makri, S. (2016). *Qualitative HCI research: Going behind the scenes*. Morgan & Claypool Publishers.
7. Boschi, V., Catricala, E., Consonni, M., Chesi, C., Moro, A., & Cappa, S. F. (2017). Connected speech in neurodegenerative language disorders: a review. *Frontiers in psychology, 8*, 269.
8. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology, 3*(2), 77-101.

9. Carruthers, H. R., Morris, J., Tarrier, N., & Whorwell, P. J. (2010). The Manchester Color Wheel: development of a novel way of identifying color choice and its validation in healthy, anxious and depressed individuals. *BMC medical research methodology, 10*(1), 12.
10. Carver, C. S. (1997). You want to measure coping but your protocol' too long: Consider the brief cope. *International journal of behavioral medicine, 4*(1), 92-100.
11. Chang, S. J., Koh, S. B., Kang, D., Kim, S. A., Kang, M. G., Lee, C. G., ... & Son, D. K. (2005). Developing an occupational stress scale for Korean employees. *Annals of occupational and environmental medicine, 17*(4), 297-317.
12. Cho, Y. W., Song, M. L., & Morin, C. M. (2014). Validation of a Korean version of the insomnia severity index. *Journal of clinical neurology (Seoul, Korea), 10*(3), 210.
13. Christensen, R. H., Bentzen, B. H., Andersen, C. M., et al. (2020). Acute effects of a high-fat meal on heart rate variability: A randomised controlled trial. Frontiers in Physiology, 11, 926.
14. Chung, H., et al. (2021). Passive Sensing in Mental Health: A Review of Ethical Concerns. npj Digital Medicine.
15. Cohen, A. S., & Elvevåg, B. (2014). Automated computerized analysis of speech in psychiatric disorders. *Current opinion in psychiatry, 27*(3), 203-209.
16. Cohen, A. S., McGovern, J. E., Dinzeo, T. J., & Covington, M. A. (2014). Speech deficits in serious mental illness: a cognitive resource issue?. *Schizophrenia research, 160*(1-3), 173-179.
17. Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech communication, 71*, 10-49.
18. Ding, Z., Chen, J., Zhong, B. L., Liu, C. L., & Liu, Z. T. (2025). Emotional stimulated speech-based assisted early diagnosis of depressive disorders using personality-enhanced deep learning. *Journal of Affective Disorders, 376*, 177-188.
19. Einsle, F., Köllner, V., Dannemann, S., & Maercker, A. (2010). Development and validation of a self-report for the assessment of adjustment disorders. *Psychology, health & medicine, 15*(5), 584-595.
20. Epstein, D. A., Ping, A., Fogarty, J., & Munson, S. A. (2015, September). A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing* (pp. 731-742).
21. Esgalhado, F., Batista, A., Vassilenko, V., Russo, S., & Ortigueira, M. (2022). Peak detection and HRV feature evaluation on ECG and PPG signals. *Symmetry, 14*(6), 1139.
22. Fagherazzi, G., Fischer, A., Ismael, M., & Despotovic, V. (2021). Voice for health: the use of vocal biomarkers from research to clinical practice. *Digital biomarkers, 5*(1), 78-88.
23. Hards, E., Orchard, F., & Reynolds, S. (2023). 'I am tired, sad and kind': self-evaluation and symptoms of depression in adolescents. *Child and Adolescent Psychiatry and Mental Health, 17*(1), 126.
24. Harte, R., Glynn, L., Rodríguez-Molinero, A., Baker, P. M., Scharf, T., Quinlan, L. R., & ÓLaighin, G. (2017). A human-centered design methodology to enhance the usability, human factors, and user experience of connected health systems: a three-phase methodology. *JMIR human factors, 4*(1), e5443.
25. Hernandez, J., McDuff, D., Fletcher, R., & Picard, R. W. (2013, March). Inside-out: Reflecting on your inner state. In *2013 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)* (pp. 324-327). IEEE.
26. Ji, J., Dong, W., Li, J., Peng, J., Feng, C., Liu, R., ... & Ma, Y. (2024). Depressive and mania mood state detection through voice as a biomarker using machine learning. *Frontiers in Neurology, 15*, 1394210.
27. Kim, H. G., Cheon, E. J., Bai, D. S., Lee, Y. H., & Koo, B. H. (2018). Stress and heart rate variability: a meta-analysis and review of the literature. *Psychiatry investigation, 15*(3), 235.
28. Kim, Y., Song, H., Jeon, Y., Oh, Y., & Lee, Y. (2022). Development and validation of a Korean affective voice database. *Phonetics and Speech Sciences, 14*(3), 77-86.

29. Kocalevent, R. D., Hinz, A., & Brähler, E. (2013). Standardization of the depression screener patient health questionnaire (PHQ-9) in the general population. *General hospital psychiatry, 35*(5), 551-555.
30. Koh, Z. H., Skues, J., & Murray, G. (2023). Digital self-report instruments for repeated measurement of mental health in the general adult population: a protocol for a systematic review. *BMJ open, 13*(1), e065162.
31. Kuhn, M. H., & McPartland, T. S. (2017). An empirical investigation of self-attitudes. In *Sociological Methods* (pp. 167-182). Routledge.
32. Lin, R. F., Leung, T. K., Liu, Y. P., & Hu, K. R. (2022, May). Disclosing critical voice features for discriminating between depression and insomnia-a preliminary study for developing a quantitative method. In *Healthcare* (Vol. 10, No. 5, p. 935). MDPI.
33. Lucas, G. M., Gratch, J., King, A., & Morency, L. P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior, 37*, 94-100.
34. Lyzwinski, L. N., Elgendi, M., & Menon, C. (2023). The use of photoplethysmography in the assessment of mental health: scoping review. *JMIR Mental Health, 10*, e40163.
35. Madanian, S., Parry, D., Adeleye, O., Poellabauer, C., Mirza, F., Mathew, S., & Schneider, S. (2022). Automatic speech emotion recognition using machine learning: digital transformation of mental health.
36. Mancone, S., Diotaiuti, P., Valente, G., Corrado, S., Bellizzi, F., Vilarino, G. T., & Andrade, A. (2023). The use of voice assistant for psychological assessment elicits empathy and engagement while maintaining good psychometric properties. *Behavioral Sciences, 13*(7), 550.
37. Mayer, R. E. (2009). *Multimedia Learning*. Cambridge University Press.
38. McDuff, D., Gontarek, S., & Picard, R. (2014, August). Remote measurement of cognitive stress via heart rate variability. In *2014 36th annual international conference of the IEEE engineering in medicine and biology society* (pp. 2957-2960). IEEE.
39. Mohr, D. C., Schueller, S. M., Montague, E., Burns, M. N., & Rashidi, P. (2014). The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. *Journal of medical Internet research, 16*(6), e146.
40. Myllymäki, T., Rusko, H., Syväoja, H., Juuti, T., Kinnunen, M. L., & Kyröläinen, H. (2012). Effects of exercise intensity and duration on nocturnal heart rate variability and sleep quality. *European journal of applied physiology, 112*(3), 801-809.
41. Nass, C. I., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human-computer relationship* (p. 9). Cambridge: MIT press.
42. Norman, D. (2013). *The Design of Everyday Things*. Basic Books.
43. Owens, A. P. (2020). The role of heart rate variability in the future of remote digital biomarkers. *Frontiers in Neuroscience, 14*, 582145.
44. Prins, A., Bovin, M. J., Smolenski, D. J., Marx, B. P., Kimerling, R., Jenkins-Guarnieri, M. A., ... & Tiet, Q. Q. (2016). The primary care PTSD screen for DSM-5 (PC-PTSD-5): development and evaluation within a veteran primary care sample. *Journal of general internal medicine, 31*(10), 1206-1211.
45. Rehman, R. Z. U., Chatterjee, M., Manyakov, N. V., Daans, M., Jackson, A., O'Brisky, A., ... & Morris, M. (2024). Assessment of physiological signals from photoplethysmography sensors compared to an electrocardiogram sensor: a validation study in daily life. *Sensors, 24*(21), 6826.
46. Robb, D. A., Lopes, J., Ahmad, M. I., McKenna, P. E., Liu, X., Lohan, K., & Hastie, H. (2023). Seeing eye to eye: trustworthy embodiment for task-based conversational agents. *Frontiers in Robotics and AI, 10*, 1234767.
47. Robin, J., Harrison, J. E., Kaufman, L. D., Rudzicz, F., Simpson, W., & Yancheva, M. (2020). Evaluation of speech-based digital biomarkers: review and recommendations. *Digital biomarkers, 4*(3), 99-108.
48. Salvi, J. (2019). Columbia-suicide severity rating scale (C-SSRS). *Emergency medicine practice, 21*(5), CD3.

49. Schmidt, M. E., Chang-Claude, J., & Steindorf, K. (2016). Heart rate variability: A potential early marker of changes in physical activity and sedentary behavior in cancer patients. *Frontiers in Public Health, 4*, 30.
50. Shevat, A., 2017. *Designing bots: Creating conversational experiences*. "O'Reilly Media, Inc."
51. Siemon, D., Ahmad, R., Harms, H., & de Vreede, T. (2022). Requirements and solution approaches to personality-adaptive conversational agents in mental health care. *Sustainability, 14*(7), 3832.
52. Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine, 166*(10), 1092-1097.
53. Stanley, J., Peake, J. M., & Buchheit, M. (2013). Cardiac parasympathetic reactivation following exercise: implications for training prescription. *Sports medicine, 43*(12), 1259-1277.
54. Tasnim, M., Ehghaghi, M., Diep, B., & Novikova, J. (2023). Depac: a corpus for depression and anxiety detection from speech. *arXiv preprint arXiv:2306.12443*.
55. Tasnim, M., Ramos, R. D., Stroulia, E., & Trejo, L. A. (2024, April). A machine-learning model for detecting depression, anxiety, and stress from speech. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 7085-7089). IEEE.
56. Thayer, J. F., & Lane, R. D. (2009). Claude Bernard and the heart rate variability: From bench to bedside and back. *Frontiers in Neuroscience, 3*, 14.
57. Torous, J., Lipschitz, J., Ng, M., & Firth, J. (2020). Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. *Journal of affective disorders, 263*, 413-419.
58. Zhuang, X., Rozgić, V., Crystal, M., & Marx, B. P. (2014, December). Improving speech-based PTSD detection via multi-view learning. In *2014 IEEE Spoken Language Technology Workshop (SLT)* (pp. 260-265). IEEE.
59. Zimmermann, J. J., Tiellet Nunes, M. L., & Fleck, M. P. (2018). How do depressed patients evaluate their quality of life? A qualitative study. *Journal of Patient-Reported Outcomes, 2*(1), 52.
60. de Zambotti, M., Nicholas, C. L., Colrain, I. M., Trinder, J. A., & Baker, F. C. (2013). Autonomic regulation across phases of the menstrual cycle and sleep stages in women with premenstrual syndrome and healthy controls. *Psychoneuroendocrinology, 38*(11), 2618-2627.