



# VoiceCraft: Designing Task-specific Voice Assistant Personas

Mateusz Dubiel  
University of Luxembourg  
Luxembourg  
mateusz.dubiel@uni.lu

Nima Zargham  
zargham@uni-bremen.de  
Digital Media Lab, University of Bremen  
Germany

Smit Desai  
University of Illinois at Urbana-Champaign  
USA  
Northeastern University  
USA  
smitad2@illinois.edu

Anuschka Schmitt  
University of St.Gallen  
Switzerland  
anuschka.schmitt@unisg.ch

## ABSTRACT

VoiceCraft workshop aims to establish a research community focused on the design and evaluation of Voice Assistant (VA) personas for both task-oriented functions (e.g., information search, online shopping) and personal growth applications (e.g., coaching, mindful reflection, tutoring). Through discussion and collaborative efforts, we will seek to propose a set of practices and standards that will help to improve the ecological validity of VA personas. In particular, we will explore topics such as the interaction design of voice-based interfaces, the impact of agent personas on the user experience, and the approaches for designing such VA agents. This workshop will serve as a platform to build a better-equipped community to explore VA personas that provide a better fit to range of everyday interaction scenarios.

## CCS CONCEPTS

- Human-centered computing → HCI design and evaluation methods.

## KEYWORDS

Design, Personas, Speech Interfaces

### ACM Reference Format:

Mateusz Dubiel, Smit Desai, Nima Zargham, and Anuschka Schmitt. 2024. VoiceCraft: Designing Task-specific Voice Assistant Personas. In *ACM Conversational User Interfaces 2024 (CUI '24)*, July 08–10, 2024, Luxembourg, Luxembourg. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3640794.3670000>

## 1 THEME AND GOALS

The theme of the workshop is “VoiceCraft: Designing Task-specific Voice Assistant Personas”. Increasingly, AI-based artifacts are embedded in interfaces that interact with the user in a dialogue-based manner, via text and voice. Indeed, voice-based user interaction with AI-based artifacts can be viewed as an emerging and essential

facet of HCI [5]. The growing popularity and increased usage of interfaces that feature synthetic speech could be partly attributed to their ever-improving natural language processing capabilities. Recent developments in Deep Learning have contributed to a rapid improvement of the quality of synthetic voices in terms of intelligibility and naturalness, making them almost indistinguishable from human speech [13, 18].

Conversational agents, including Amazon Alexa and Google Home, are adept at embodying complex personas, such as financial advisors or exercise coaches. These agents facilitate various everyday tasks that have significant implications for users’ finances—such as ordering takeaways or making online purchases [10]. Additionally, they serve in roles like instructors or health coaches, where they offer valuable health-related information [8] and tailored exercise programs [9], further demonstrating their versatility and impact in daily life. Despite rapid advancements in speech technology and its widespread commercial deployment, research into the design of these voices and their implications for users remains limited [16]. This scarcity of studies contributes to inconsistencies in the development and evaluation of system personas—such as teachers [7] or therapists [15]—that utilize synthetic voices. These personas are often assessed through interactive user studies, which occur both online and in laboratory settings, yet lack uniform standards and methodologies.

To address these challenges, this workshop will convene interdisciplinary experts in audio engineering, speech perception, UX design, and related fields to refine how voice user interfaces are crafted, specifically focusing on the development of persona-specific voice assistants. The goal is to enhance design practices and increase the ecological validity of these systems. Throughout the workshop, through interactive demonstrations, vibrant group discussions, and hands-on prototyping activities, we will explore key questions tailored to the creation of effective and engaging personas, such as:

- *How to choose and design appropriate VA personas based on the context of use?*
- *How to effectively tailor VA personas’ voices to different application domains?*
- *What are the sociotechnical and ethical implications of designing VA personas?*

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

*CUI '24, July 08–10, 2024, Luxembourg, Luxembourg*

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0511-3/24/07

<https://doi.org/10.1145/3640794.3670000>

We encourage collaboration between academia and industry and welcome participation from anyone interested in conversational user interface design, development, and evaluation. By bringing the interdisciplinary community together, our goal is to provide insights on how to improve development of VA personas that are tailored to specific applications, and ensure higher ecological validity of their task-based evaluation.

## 2 BACKGROUND AND MOTIVATION

VAs lack visual affordances and, thus, unlike Graphical User Interfaces (GUIs), rely on ‘personas’ to act as the system’s front end. Personas are personalities assigned to VAs to establish familiarity with the users and form a crucial part of the VA design process [4]. In previous literature, VAs have been designed to occupy various personas, such as instructors, teachers, therapists, and coaches [6]. Designing personas involves choosing personality traits or adjectives, imagining a biography, and selecting or designing an appropriate voice to interact with the end-user.

Since auditory cues such as pitch variance or speaking rate and their modifications can affect user perception and behaviour (e.g., [1, 11, 12, 14, 17]), exploring such cues will be the focus of our workshop. Previous work has demonstrated that modifying the voice characteristics of a speech agent can significantly influence user experience. For instance, Zargham et al. [19] introduced the concept of multi-agent voice assistants, employing multiple synthetic voices to embody various personas within a single voice user interface. Their findings indicated that users distinguished each voice as a unique character with a distinct personality, expressing a preference for this multi-agent approach over conventional single-agent systems. Participants perceived different voices as distinct characters with personalities. They believed certain agents could be more suitable for specific task domains, conveying different character traits through voice factors like tone, gender, or accent.

Voicecraft is a venue that will bring audio engineers, UX designers, and speech perception experts together to discuss how to design and implement VA personas that are engaging and provide good fit to a particular task.

## 3 ORGANISERS

**Mateusz Dubiel** is a Research Associate in the Department of Computer Science at the University of Luxembourg, where he works on development and evaluation of conversational agents. Specifically, his current research focuses on assessment of cognitive and usability implications of interfaces that feature speech, and exploration of their potential to inspire positive behavioural change in users. He served as Short Papers Chair for CUI '22 and is one of General Chairs for CUI '24

**Smit Desai** is a Ph.D. candidate in the School of Information Sciences at the University of Illinois, Urbana-Champaign. His primary research focus centers around comprehending the mental models of users as they engage with conversational agents, utilizing innovative research techniques such as metaphor analysis. He leverages this valuable insight to advance the development of conversational agents in diverse social roles, including educators and storytellers. His research has yielded publications in esteemed HCI forums like CHI, TOCHI, CSCW, and CUI.

**Nima Zargham** is a Ph.D. student in the Digital Media Lab at the University of Bremen. His research focuses on human-centered approaches for designing speech-based systems that elicit desirable user experiences. Nima has previously organized CUI-related workshops at notable conferences such as ACM/IEEE HRI 2023, ACM CUI 2023, and ACM CHI 24. Additionally, he served as a local chair at the ACM CHI-PLAY 2022 conference. His research efforts have resulted in publications featured in prestigious HCI venues, including CHI, CUI, and CHI-PLAY.

**Anuschka Schmitt** is a research associate and Ph.D. candidate at the Institute of Information Management at the University of St.Gallen (HSG), Switzerland and currently a visiting researcher at Harvard University’s School of Engineering and Applied Sciences. Her research focuses on user trust and decision-making in human-computer interaction, as well as perceptual and behavioural implications of conversational AI.

## 4 SCHEDULE AND DESCRIPTION OF ACTIVITIES PLANNED

The tentative schedule and workshop activities are presented in Table 1. Since the purpose of our workshop is to synthesise insights from experts from different communities and translate them into practical persona design and evaluation recommendations, we will ensure that ample time is dedicated to prototyping activities, discussions, and interactions between the participants.

**Table 1: Tentative Workshop Plan**

Time Slot (duration)	Activity
09:00-09:15 (15mins)	Welcome and Introductions
09:15-10:15 (60mins)	Keynote by Prof. Matthew Aylett
10:15-10:30 (15mins)	Coffee Break
10:30-10:45 (15mins)	Expressive TTS Demonstration
10:45-11:30 (45mins)	Group Prototyping Session
11:30-12:00 (30mins)	Group Reporting and Discussion
12:00-12:30 (30mins)	Closing Remarks and Future Plans

In order to inspire participants and spur discussions, the workshop will begin with a keynote delivered by Prof. Matthew Aylett<sup>1</sup>, who will explain how drawing inspiration from human dramatic performance can help us design interactive speech synthesis systems that can be finely controlled by a human voice. The first part of the workshop will also include a demo session, where participants will be presented with examples of anthropomorphic voice agents, followed by a discussion about their personas.

The demonstration will be followed by the group prototyping session, where participants will work together to design their own persona for a specific application. To foster interdisciplinary collaboration, each group will include participants from different backgrounds. In order to facilitate this process and assist participants with any questions, each group will be moderated by at least one workshop organiser. The participants will have 45 minutes to design their personas and prepare short presentations motivating the design process and proposing an evaluation criteria. Then after

<sup>1</sup><https://www.edinburgh-robotics.org/academics/matthew-aylett>

each group will give a short presentation showcasing their persona and explaining their fit to the proposed interactive scenario (30 minutes). This followed by 30 minutes of brainstorming activity and summarising the insights. The exact format and duration of group reporting and discussion will be based on the final number of workshop attendees (max. 20 people).

## 5 ANTICIPATED OUTCOMES

We anticipate that our workshop will provide both theoretical and practical outcomes. As for theory, the workshop will: (1) bring a better, cross-disciplinary understanding of how VA personas should be developed to provide the best fit for the designated task and ensure the best user experience, and (2) help to inform a set of standardised evaluation criteria to ensure that systems that feature such personas are evaluated in a more ecologically valid way.

In terms of practical outcomes, the 'Voicraft: Designing Task-specific Voice Assistant Personas' workshop will lead to the formation of a new research community comprised of academics, industry experts and members of the public that will promote research on the interactive, task-based evaluation of VA personas and continuously identify and tackle new challenges as they arise. Ultimately, we anticipate that the concerted efforts of our newly formed community will lead to higher robustness and standardisation of VA personas design and evaluation. In summary, the workshop will lead to:

- (1) Fostering a closer collaboration between the CUI community members and speech perception researchers.
- (2) Synthesising views from different areas of expertise (i.e., audio engineering, UX design and speech perception).
- (3) Better design and development practice, and improved ecological validity of VA evaluation studies.

## 6 RESULTS DISSEMINATION BEYOND CUI 2024

The workshop outcomes will be presented as a report and posted on the workshop's website to reach a broader audience. The organisers will use social media platforms (Mastodon and X (formerly Twitter)) to disseminate these materials.

## ACKNOWLEDGMENTS

This work is in part supported by European Innovation Council Pathfinder program (SYMBIOTIK project, grant 101071147).

## REFERENCES

- [1] Pascal Belin, Bibi Boehme, and Phil McAleer. 2017. The sound of trustworthiness: Acoustic-based modulation of perceived voice personality. *PLoS one* 12, 10 (2017), e0185651.
- [2] Julia Cambre and Chinmay Kulkarni. 2019. One voice fits all? Social implications and research challenges of designing voices for smart devices. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–19.
- [3] Vijay Chidambaram, Yueh-Hsuan Chiang, and Bilge Mutlu. 2012. Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. 293–300.
- [4] M. Sadek, R. A. Calvo, and C. Mougenot. Trends, Challenges and Processes in Conversational Agent Design: Exploring Practitioners' Views through Semi-Structured Interviews. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (CUI '23), pages 1–10, New York, NY, USA, July 2023. Association for Computing Machinery.
- [5] Leigh Clark, Philip Doyle, Diego Garaialde, Emer Gilmartin, Stephan Schlögl, Jens Edlund, Matthew Aylett, João Cabral, Cosmin Munteanu, Justin Edwards, et al. 2019. The state of speech in HCI: Trends, themes and challenges. *Interacting with computers* 31, 4 (2019), 349–371.
- [6] Smit Desai and Michael Twidle. Metaphors in Voice User Interfaces: A Slippery Fish. *ACM Transactions on Computer-Human Interaction*, 30(6):89:1–89:37, September 2023.
- [7] Smit Desai and Jessie Chin. Hey Google, Can You Help Me Learn? In *Proceedings of the 3rd Conference on Conversational User Interfaces (CUI '21)*, Bilbao (online), Spain, 2021. Association for Computing Machinery. DOI: [10.1145/3469505.3469601](https://doi.org/10.1145/3469505.3469601). ISBN: 9781450389983. Article No.: 6, Number of pages: 4. Keywords: Computer-Based Learning Environments, Informal Learning, Metacognition, Self-Regulated Learning, Voice User Interface.
- [8] Smit Desai and Jessie Chin. OK Google, Let's Learn: Using Voice User Interfaces for Informal Self-Regulated Learning of Health Topics among Younger and Older Adults. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*, Hamburg, Germany, 2023. Association for Computing Machinery. DOI: [10.1145/3544548.3581507](https://doi.org/10.1145/3544548.3581507). ISBN: 9781450394215. Article No.: 847, Number of pages: 21. Keywords: Aging, Conversational agent(s), Health information, Informal Learning, Interaction fluency, Metacognition, Self-regulated learning, Technology acceptance, Voice assistant(s), Voice user interface(s).
- [9] Smit Desai, Xinhui Hu, Morgan Lundy, and Jessie Chin. Using Experience-Based Participatory Approach to Design Interactive Voice User Interfaces for Delivering Physical Activity Programs with Older Adults. In *Proceedings of the 11th International Conference on Human-Agent Interaction (HAI '23)*, pages 180–190, Gothenburg, Sweden, 2023. Association for Computing Machinery. DOI: [10.1145/3623809.3623827](https://doi.org/10.1145/3623809.3623827). ISBN: 9798400708244. Keywords: Co-design, Conversational agents, Older Adults, Participatory Design, Physical Activity, Voice User Interfaces.
- [10] Mateusz Dubiel, Sylvain Daronnat, and Luis A Leiva. 2022. Conversational Agents Trust Calibration: A User-Centred Perspective to Design. In *Proceedings of the 4th Conference on Conversational User Interfaces*. 1–6.
- [11] Mateusz Dubiel, Martin Halvey, Pilar Oplustil Gallegos, and Simon King. 2020. Persuasive synthetic speech: Voice perception and user behaviour. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. 1–9.
- [12] Aaron C Elkins and Douglas C Derrick. 2013. The sound of trust: voice as a measurement of trust during interactions with embodied conversational agents. *Group decision and negotiation* 22, 5 (2013), 897–913.
- [13] Andrew Gibiansky, Sercan Arik, Gregory Diamos, John Miller, Kainan Peng, Wei Ping, Jonathan Raiman, and Yanqi Zhou. 2017. Deep voice 2: Multi-speaker neural text-to-speech. *Advances in neural information processing systems* 30 (2017).
- [14] Phil McAleer, Alexander Todorov, and Pascal Belin. 2014. How do you say 'Hello'? Personality impressions from brief novel voices. *PLoS one* 9, 3 (2014), e90779.
- [15] Nasim Motalebi, Eugene Cho, S. Shyam Sundar, and Saeed Abdullah. Can Alexa be your Therapist? How Back-Channeling Transforms Smart-Speakers to be Active Listeners. In *Companion Publication of the 2019 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '19 Companion)*, pages 309–313, Austin, TX, USA, 2019. Association for Computing Machinery. DOI: [10.1145/3311957.3359502](https://doi.org/10.1145/3311957.3359502). ISBN: 9781450366922. Keywords: wellbeing, smart-speakers, conversational agents, back-channeling, active listening.
- [16] Anuschka Schmitt, Naim Zierau, Andreas Janson, and Jan Marco Leimeister. 2021. Voice as a contemporary frontier of interaction design. In *European Conference on Information Systems (ECIS)-Virtual*.
- [17] Mateusz Dubiel, Anastasia Sergeeva, and Luis A Leiva. 2024. Impact of Voice Fidelity on Decision Making: A Potential Dark Pattern?. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*. 181–194.
- [18] Klie, L. AI Breathes Life into an Industry on the Edge.. *Speech Technology Magazine*. 29, 4–5 (2024).
- [19] Zargham, N., Bonfert, M., Porzel, R., Doring, T. & Malaka, R. Multi-Agent Voice Assistants: An Investigation of User Experience. *Proceedings Of The 20th International Conference On Mobile And Ubiquitous Multimedia*. pp. 98–107 (2021)