Designing for Radar Pace: A Conversational System for Coaching

Abstract
As the success of Siri, Alex, and the Google Assistant have demonstrated, conversational agents are becoming more and more commonplace. With the proliferation of this interaction modality, new methods and associated design challenges are emerging. In some cases, we can draw on lessons learned in designing interactive voice response (IVR) systems or from human-to-human communication paradigms. In other cases, we need new ways to deal with the design challenges created by the current state of the technology. In this paper, we present Radar Pace, a conversational coach for running and cycling developed by Oakley and Intel, and highlight 5 design challenges that were tackled during the course of developing the product.

Introduction
Radar Pace (see Figure 1 for a photo of the product) is a conversational coaching system that provides real-time feedback for running and cycling. The product was
developed as part of a partnership between Oakley and Intel and launched in 5 languages (English, Spanish, German, Italian and French) in October 2016. The user was able to talk to the system via a microphone embedded in the stem of the glasses, and the system responded via custom Text-to-Speech (TTS) voices through the earbuds. Each language had its own voice, gender and personality based on the country it was targeted to launch in.

The expert coaching system and associated dialogue system powering the conversational UI were packaged up as part of a companion mobile application, launched for both Android and iOS.

During the development process, there were several design choices that made the experience more pleasant and natural for users. In this paper, we go over five of the most important design problems addressed:

1. How to naturally engage with the system by allowing the user to easily ask follow up questions, interrupt the system, and reduce the need for using a trigger phrase.
2. How to communicate to the user when the system was listening and when it heard what the user was asking.
3. How to increase discoverability of features the user could ask the system about and how to communicate the dialogue domain boundaries.
4. How to communicate why the system didn't understand the user.
5. How to design the personality, including personality traits, gender, branding, and level of formality.

We’ll go over each of these design choices in the sections below and provide a brief overview of the overall usage of Radar Pace.

**Open Mic. vs. Trigger Phrase**

While an open microphone interaction is the most natural for people (we don’t repeat a person’s name every time we say something new to them), implementing a full open mic. system has disadvantages: it consumes a lot of battery power, and makes it more likely the automatic speech recognition (ASR) system will recognize cross talk or other noise as speech and respond inappropriately. Because of these constraints, we implemented a trigger phrase, but also kept the microphone open in certain circumstances.

The conversational coach was designed as a mixed-initiative system in which the coach would initiate conversations with the user when it had relevant contextual feedback and provide general tips to the user in low cognitive load situations. In cases in which the coach initiated a conversation with the user, the system would keep listening for 30 seconds after in case the user wanted to ask multiple questions in a row or had a follow up to the original answer they received (e.g. User: “What’s my cadence?”, Radar: “88”, User: “Is that okay?” Radar: “Your cadence is just right.”).

The user could also engage with the system at any other time by saying a trigger phrase (“Ok Radar”). Again, once the user initiated the conversation, the system would continue listening for 30 seconds in case the user wanted to ask multiple questions in a row or had a follow up to the original answer they received (e.g. User: “What’s my cadence?”, Radar: “88”, User: “Is that okay?” Radar: “Your cadence is just right.”).
Figuring out how long to keep the microphone open required testing out the system with different lengths of time with users. If the window was too short, the user didn’t have time in between working hard and breathing to ask any follow up questions. If it was too long, the system would likely pick up cross-talk, since users tend to go out running or riding with others.

**Communicating Listening and Speaking States**

Because we implemented a mixed-initiative system that kept the microphone open sometimes and required the use of a trigger phrase other times, we needed to clearly communicate to the user whether or not the system was listening at any given moment. We did this through the use of tones. This also helped improve perceived responsiveness of the system, since passing audio back and forth via Bluetooth from the wearable to the phone application incurred lag that may otherwise have been unacceptable. We used tones to communicate:

- When the trigger phrase was recognized and the audio channel was opened.
- When the user’s utterance was recognized.
- When the audio channel closed.

A tone was not played when the coach spoke, since users didn’t seem to require a tone before hearing the coaching output. Without the tones, users didn’t know if...
the system was listening or not. We also learned that users can sometimes miss the tones, especially in high noise environments.

**Increasing Discoverability**

One of the primary issues with voice assistants is that communicating to the user what the voice assistant can do for them is difficult and discoverability is low: users don’t know where to start or what to ask. For Alexa, Amazon resolves this by sending out weekly “What’s new with Alexa?” emails that provide users a list of new things they can ask the system.

For Radar Pace, we created an out of box experience, where users were guided through a mini tutorial and provided with different things they could ask the system. We kept this short, and required first-time users to put on the glasses and go through every one of these screens, although we didn’t require that they actually ask the questions (see Figure 2). Users were likely to skip the tutorial if they were allowed to, which resulted in users that were not sure what they could ask the coach via voice.

Even though users weren’t required to actually ask the questions that were presented to them, we found that requiring them to go through the screens and put on the glasses meant that most users asked the system the questions presented to them anyway.

**Managing High Noise Environments**

Because the product was targeted for both running and cycling, it needed to work in high noise environments, such as cycling at 25 mph with wind noise. Managing user expectations in this context was difficult and we explored several ways of doing so. One of the ways was to use the signal-to-noise ratio to determine whether we should respond specifically to the user request with “It’s too loud, please ask again later.” An alternative, and the one that was implemented in the final product was to train the ASR system to ignore anything that didn’t seem like user speech. As a result, the system would ignore speech that it couldn’t understand in high noise environments. This resulted in a more reliable experience, as we discovered that a lot of non-speech noise was originally triggering responses from the conversational interface.

**Personality**

Designing the personality required us to understand Oakley’s brand, the product’s target audience, and any social norms or expectations for the countries we were launching in, both in relation to voice assistants in general and specifically in the coaching domain.

We conducted research in every country to understand user expectations around coaching and drew on existing general cultural research provided by [1]. Based on this, we selected an informal register for every language. French, German and Italian had male coaching voices, and English and Spanish had female ones. Oakley wanted all coaching personalities to be supportive (not like drill sergeant), but the level of support provided varied by language based on cultural expectations. For a more detailed overview of the Personality design process, see [2].

**Usage Overview**

Overall, we found that users were using the conversational UI quite a bit (see Figure 3 and Figure 4) and that usage normalized over time (see Figure 5). What this means is that users that didn’t ask many or
any questions at the beginning slowly started to ask questions as they had more and more workouts. Users that started out over-eager and asked many questions at the beginning, slowly settled on asking fewer questions over time.

Figure 3: Nearly all dialogue sessions contained some interaction with the conversational UI.

Figure 4: Generally, there was a linear increase in the number of user utterances with length of workout.

Figure 5: Number of user inputs normalized to between 8-30 over time.

Finally, questions about the user’s plan, their metrics, and requests for help were the features that were most used in Radar Pace.

References