MUSE: A VOICE-ACTIVATED MUSIC SEARCH AGENT FOR THE *DIAL*

Yuan-Yi Fan
I.AM+, LLC
danny.fan@iamplus.com

Vids Samanta
I.AM+, LLC
vids.samanta@iamplus.com

ABSTRACT

MuSe is a software agent that enables music discovery via voice on the *dial* (https://iamplus.com/dial/). The goal of this demo is to make audio-centric wearable user experience accessible to a broader audience. This demo will focus on voice-enabled music content access and discovery features on the *dial*, where people can interact with our voice assistant AneedA and ask her music-related queries, such as "tell me more about this artist", "show me similar artists", "show me similar songs", "when is the next show", "what's up with [artist]", and "play [song] by [artist]".

1. INTRODUCTION

Voice assistant technology, online music streaming services, and music metadata providers have expanded the design space for audio-centric wearable user experience (UX). To navigate this emerging design space, we developed MuSe, a Natural Language Understanding (NLU) agent for Music Information Retrieval (MIR) tasks on the *dial*. In addition to applying auditory display concepts [1] to the music experience design on the *dial*, we've identified that NLU and MIR are two essential technologies to advance audio-centric UX. In our opinion, further research that draws on both technologies could enhance musical intelligence of the voice assistant and enable deeper conversations on the *dial*.

2. BACKGROUND

With an Amazon Echo speaker, you can control a music player with voice commands, and even ask the voice assistant "Alexa, buy this song". However, the "Alexa" wakeup keyword with the QA-style conversation make the UX less natural. In the wireless speaker market, SONOS is looking to the future in streaming services and voice control. [2] Similar to Amazon's Alexa Skill Kit, Soundhoud's Houndify has recently opened their developer platform aiming to add voice-enabled conversational interface to anything. [3] The recent collaboration between Soundhound and NVIDIA on bringing deep learning-based NLU to in-car infotainment system also suggests the increasing significance of a more intelligent voice assistant in enhancing the audio-centric UX. [4] Another strong indication of this emerging audio-centric UX is the collaboration between Gracenote and Nuance. Starting from 2012, the 10-year strategic partnership between Gracenote and Nuance aims to enable users to discover music and videos via voice, across all kinds of platforms, from mobile phones, tablets, TVs, to cars. [5] VoCon Music Premium, as the first product of their partnership, allows users to query artist information via voice commands. Recently, Google released a voice-activated home product capable of streaming music and managing everyday tasks. It remains to be seen what Google plans for their assistant on Google Home. [6]

Using natural language utterances for music retrieval and discovery is not new [1–3]. However, the collected references in this section demonstrate growing interests in this emerging design space of audio-centric UX.

3. DEMO DESCRIPTION

To initiate a conversation with the *dial*, a user starts by pushing the triangle button by the display. Our MuSe agent enables the voice assistant AneedA on the *dial* to respond to the user's music-related queries via text-to-speech with a synchronized visual UI.

Figure 1. The Dial.

[1] Text-to-speech, Audio Focus, Earcons.

[2] http://venturebeat.com/2016/03/09/sonos-announces-layoffs-looks-to-future-in-streaming-services-and-voice-control/

[3] https://www.houndify.com

[4] http://www.businesswire.com/news/home/20160105006247/en/SoundHound-Collaborates-NVIDIA-Bring-Deep-Learning-Based-Natural

[5] http://www.nuance.com/company/news-room/press-releases/nuangracenoteweb.doc

[6] https://home.google.com/

We launched the *dial* product in the UK in May 2016. The *dial* has a built-in microphone and speaker, and can also connect to bluetooth earphones. This demo will only focus on the voice-enabled music content access and discovery features on the *dial*. The supported use cases and voice commands for music content access and discovery are listed in the following subsections.

3.1 Use case 1: retrieving artist information

A user can get biographical information about an artist or band by asking AneedA questions like "who is this artist?", "where are they from?", and "who are the band members?". AneedA will speak out a short biography of the requested artist and show an artist image on the display.

3.2 Use case 2: discovering similar music and artists

A user can ask AneedA to show similar artists or bands to discover similar artists and bands. AneedA will speak out the names of five similar bands and show their images on the screen. Tapping on any one of these will allow the user to see a biography of the related artist or play their songs. A user could also ask AneedA for songs related to the song that is currently playing.

3.3 Use case 3: retrieving music content

This unique feature allows a user to retrieve music via a voice command with a qualifier. For example, a user can press the AneedA button and ask "Play The Black Eyed Peas before Fergie."

3.4 Use case 4: creating playlists

Use case 3 and 4 usually happen in the same conversational context. A user can add a song to a playlist by giving a follow-up command in the same conversational context.

User: Play Toxic.
(The *dial* playing Toxic)
User: Add this to my workout collection.
AneedA: You don't have a workout collection, should I create one?
User: Yes, please.
AneedA: Ok, added Toxic to your workout collection.

3.5 Use case 5: identifying music

To identify a song, a user can press the AneedA button and ask "what song is playing?".

User: What's playing?
AneedA: This is "Down in the DM" by Yo Gotti.
User: Add this to my favorites.
AneedA: Dope! That's a great song. Added it to your favorites.

3.6 Use case 6: retrieving concert information

A user can also ask AneedA about upcoming concerts for a band and AneedA will present concert information back to the user as seen in the screenshot below. As you can see in this example the system is context aware and so when the user doesn't specify the name of the artist or band the system infers the artist name for the currently playing track.

Figure 2. Request

Figure 3. Response [7]

3.7 Use case 7: retrieving the latest news about artist

To retrieve the latest information about an artist, a user can press the AneedA button and ask "what's up with [artist]?". AneedA will speak out five latest news titles about the requested artist.

Figure 4. Request

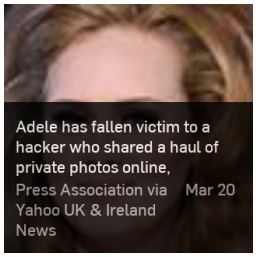

Figure 5. Response

4. ACKNOWLEDGEMENT

The authors thank Will.i.am, Chandrasekar Rathakrishnan, Duncan Burns, TVS Deepak, and Pooja Kushalappa for their support in developing MuSe.

5. REFERENCES

[1] Stephan Baumann and Andreas Klüter. Superconvenience for non-musicans: Querying mp3 and the semantic web. In *Proc. of the 3rd International Conference on Music Information Retrieval (ISMIR*, 2002.

[2] Brian McFee and Gert RG Lanckriet. The natural language of playlists. In *Proc. of the 12th International Conference on Music Information Retrieval (ISMIR*, pages 537–542, 2011.

[3] Björn Schuller, Gerhard Rigoll, and Manfred Lang. Multimodal music retrieval for large databases. In *Multimedia and Expo, 2004. ICME'04. 2004 IEEE International Conference on*, pages 755–758, 2004.

[7] Image credit: Bandsintown, LLC