INTRODUCTION

Welcome in Te Reo Māori

Tēnā koutou katoa.
Ko Barwon te awa. Ko Kathy Reid tōku ingoa.
Ngā mihi nui ki a koutou!
Kia kaha, te reo.

Welcome in English

Greetings everyone. My name is Kathy Reid, I'm from Geelong, by the Barwon River.

Thank you so much for joining this talk today, and thank you Arjen for the invitation.

Background

Up until recently, I led Developer Relations at Mycroft.AI, an open source voice assistant startup based in Kansas City, in Missouri. Mycroft are quite unique in that they provide an end to end open source voice assistant stack.

What I'd like to walk through today is an introduction to the voice stack, using the MycroftAI open source voice tools.

The reason that I think this is important to Education and to the Open Education Miniconf is that over time, voice interfaces are become more prominent - just as we moved away from command lines to graphical user interfaces, voice user interfaces are becoming more common. This means that our students are going to need an understanding of how they work, how to work with them, and how to design for them.

Questions

I'll be delighted to answer your questions, but please do save them up until the end of the talk.

OVERVIEW

Introduction to the voice stack

So, what we're going to be doing today is walking through the voice stack. Just like a web service, or an application service, a voice service has multiple layers in the stack - so the concept will be very familiar to you.

Many of you in the room I suspect are educators, or work in the education space - is that correct?

Check room

Great! So my intention here is to provide an introduction, and to showcase some of the open source technologies that are available in voice, in case you're thinking about including them in your curriculum or as part of a digital technologies program.

How many people in the audience are familiar with some type of speech recognition or text to speech service? count how many OK, that's great!

OK, that's great!

So, here's an anatomy of a voice interaction.

They're described in this diagram, by my Mycroft AI colleague, David Smehlik.

IMAGE: Anatomy of a voice interaction

In a voice interaction you begin with a **Wake Word** - also called a **Hot Word** - which is used to prepare the voice assistant to receive a command. Next, we use a **Speech to Text** engine to transcribe an **Utterance** from voice sounds into written language. Next, we use an **Intent parser** to determine what *type* of command the user wanted to execute. Then, we select a command to run and execute it. Next, we turn written language back into voice sounds again using a **Text to Speech** engine.

Does that make sense? OK, great.

Now, I'm going to dive a little deeper into each of the different layers of the voice stack, and I'm going to use the Mycroft platform to demonstrate this.

WAKE WORD

SLIDE: Wake Word summary

A Wake Word is also called a hot word and this is what a voice system uses to know that it's about to receive a command.

- always listening
- accuracy depends on what it's trained on

Able to change your Wake Word - using home.mycroft.ai

SPEECH TO TEXT

SLIDE: Speech to Text

Accurate Speech to Text conversion is one of the most challenging parts of the open source voice stack.

INTENT PARSERS

SLIDE: Intent Parsers

A strong voice stack also needs to ensure that the **intent** of the user is accurately captured. There are several open source intent parsers available.

TEXT TO SPEECH

SLIDE: Text to Speech

DEMO: Mimic recording studio

CONCLUSION

SLIDE: Thank you