Article

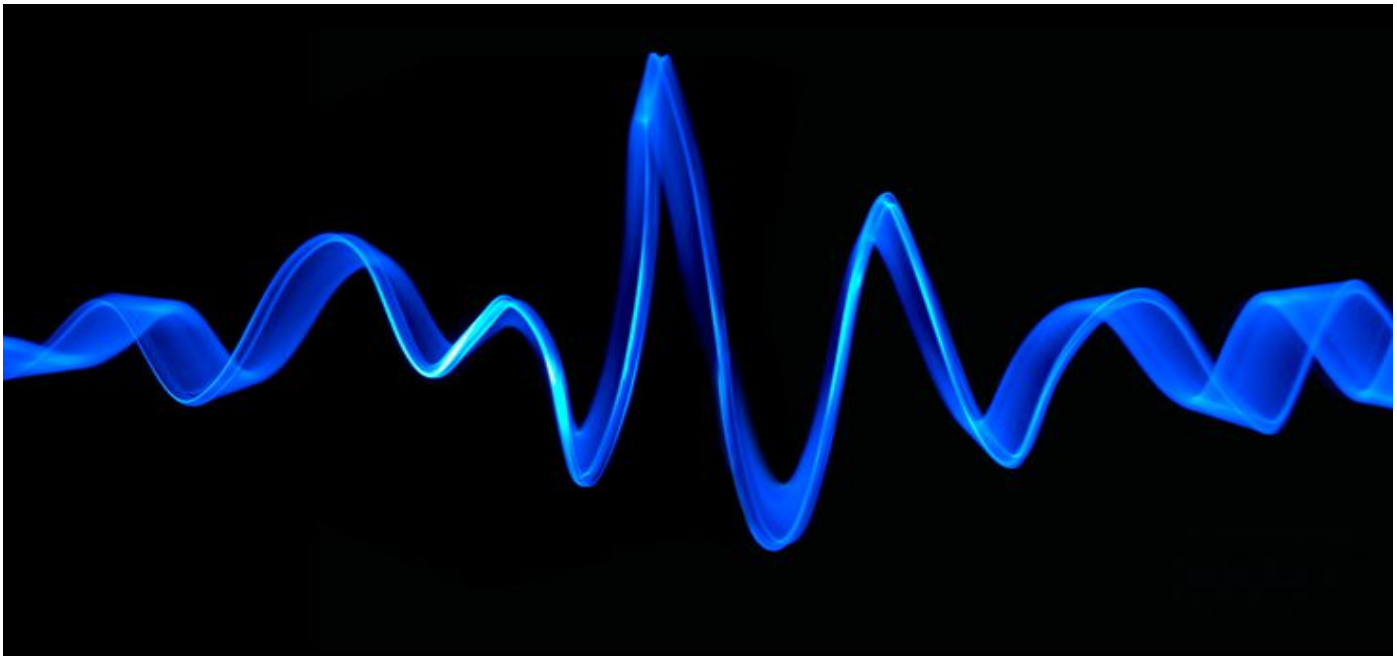
Apr 2022

# Can a machine really sound like me?

As phishing attacks become more advanced and commonplace, it is more important than ever to be able to defend yourself against them.

Daniel Halliday
Principal Consultant - Special Projects

Email daniel.halliday@waterstons.com
Linkedin https://www.linkedin.com/in/d-halliday/



## Written by Daniel Halliday and Andrew Blance.

## AI-enabled attacks

It may seem like AI (artificial intelligence)-enabled attacks are something from the future, but in reality, they are much more likely in this era than you may think.

We typically believe phishing attacks to be emails claiming to be someone asking for passwords or bank details; we've become good at spotting them and understand (hopefully) not to give important information to untrusted and suspicious sources. However, what if this email got followed up with a call from your manager, or your CEO asking for you to reply? What would you do then?

AI's ability to impersonate someone is well-documented; celebrities and politicians are regularly "deepfaked" (where videos can be made of them saying and doing things they have never done using AI). Since we believe only high-profile individuals could be victims of this sort of technology, it is easy to think that it does not apply to us; a risky attitude to have as. it is becoming easier to produce a realistic impersonation that is capable of fooling any of us in the correct context. This form of AI-enabled phishing attack may well be popular very shortly.

**Would you be fooled?**

Below is an audio file of Daniel introducing this article - but there is a twist! The recording has been fed through an AI model designed to synthesise speech – to learn to talk like a given target. This audio file is not only Daniel talking, but instead parts are also spoken by a computer. Can you tell which parts are human and which are machine?

**Reveal answer**

> This is **an article** about voice synthesis. Do you know, when you're **listening to someone what is real** and what is fake? Can you tell **which parts of** this recording were made by an AI?

*The parts highlighted in bold were generated by an AI which has been trained to sound like Daniel. Could you tell?*

Even if you know the person, it can be difficult to distinguish between what is real, and what is fake. Over the last few years, the technology used to develop these audio samples has rapidly advanced; audio synthesis has gone from research in universities and large tech companies, to something determined enthusiasts can do on their own.

Within Waterstons, our team have been training models to talk and experimenting with the technology since its infancy and have seen complicated parts of the process be automated away in easily available commercial products, allowing anyone with the time and some audio of a person's voice to begin synthesising new sounds.

# Phishing Attacks

Imagine you get an email claiming to be from your boss asking you to send your bank details, some passwords, or gift vouchers. On its own, it is likely you may ignore this, correctly identifying it as a phishing attack. However, what if it first you received this voicemail:

**EXAMPLE 2**

*Could you tell if it was really your boss or not? Compare it to this to the real Daniel saying the same words:*

This extra context, an attempt to give the phishing email more credibility, in the voice of your manager or someone senior in the company, can hugely sway how you perceive the phishing email. After all, to the listener, it sounds like you, so as far as they're concerned, you just tried to call them. If you received this voicemail alongside an email, would you be tricked by the phishing attempt?

This is all possible because AI can be trained to learn how people speak and how natural voice sounds. By training on thousands of voices, complex models can be created to get better and better at understanding how people generally talk, then used to 'read' some text. When it does it will try to do it in a way that emulates how it thinks a person would sound – but if you feed it audio of a specific person, you can then ask it to "speak" how they speak meaning the output will be the AI's interpretation of their voice.

The process of training large models typically takes a very long time and requires expensive, specialised equipment, but now pre-trained models exist that allow you to start your project with an AI that already has some understanding of human speech, reducing the amount of time you will need to train for.

Furthermore, there are cloud-based services that mean you no longer need to own hardware. Basically, with 20 minutes of transcribed and cleaned audio, you can generate new speech.

**Reputational Damage**

Aside from phishing attacks, there is huge potential for using this technology for reputational damage. Here's Daniel saying something positive and uncontroversial:

**Example 3**

**Original Audio:**

*"I think it's important that we invest in sustainable technology in the next year to help the environment."*

We can then use the model we have trained to add in some words, totally changing the context of his words:

**Manipulated Audio:**

Reveal Answer

> *"I think **we don't need to** invest in sustainable technology in the next year**, it's not our responsibility** to help the environment."*

This form of manipulation is very dangerous with the potential for significant reputational damage.

Changing sentences slightly to change their meaning entirely can create audio that sound more 'realistic' than sentences entirely generated by a computer. There are even more advanced models in the market that can be trained from only seconds of audio, or ones that allow you to change individual parts of words. These models could also be used to detect how similar two people sound, or even be repurposed to tell you if the person on the phone is real or not.

**So, what does that mean for my business?**

Determined enthusiasts with the time and inclination, can synthesise audio files that sound like you which could be used to falsely accuse you of saying something incriminating or legitimise a phishing attack.

If you were only to receive an email it can be easy to see it as a scam – but the additional layer of legitimacy that a phone call from your boss can give it that makes this form of phishing more dangerous.

If you want to learn more about cyber security and how it can be used to protect your business, speak to our cyber resilience team here.

To find out more about how you can utilise data and machine learning, speak to the team here.