UWE Bristol | University of the West of England

Presented by:

Andrew McCarthy

Postgraduate Researcher

Computer Science Research Centre

M.Sc. Cyber Security

PGCert Computing

B.Sc. Computing for Real-time Systems

June 17 2020
Cyber Science 2020

Shouting Through Letterboxes: A study on attack susceptibility of voice assistants

Twitter: @AndyMcCarthyUK

Authors: Andrew McCarthy, Phil Legg, Benedict Gaster
Andrew6.McCarthy@uwe.ac.uk ; Phil.Legg@uwe.ac.uk ; Benedict.Gaster@uwe.ac.uk

Author Profiles:

https://people.uwe.ac.uk/Person/Andrew6Mccarthy ; https://people.uwe.ac.uk/Person/PhilLegg
https://people.uwe.ac.uk/Person/BenedictGaster

Introduction & Motivation

- Household names: Alexa, Google, Cortana
- Smart Speakers
- Phones & Tablets
- Traditional Desktops and Laptops

Many of us really enjoy the convenience of using such assistants; we have focussed on the benefits of these systems without fully considering their vulnerabilities.

Background

- Voice is becoming more common for casual users. Alepis and Patsakis[16] argue that voice assistants are replacing traditional user interfaces.
- Users should decide on how they balance the trade off between convenience and operation with the issues of security and privacy[14].

- Worryingly accurate and easy to produce voice impersonation is possible using software systems such as VoCo[17].
- A British energy firm were tricked into paying £200,000 to fraudsters using AI software that accurately mimicked the executive's accent and style of speaking [18].
- Adversarial Attacks have been shown on facial recognition[10], road sign recognition [11], and network intrusion detection[12]

- Recent research shows against voice-controlled systems.
- The adversarial attack CommanderSong [23] inserts voice commands into music videos or audio files with minor perturbations, resulting in normal sounding audio; however voice assistants recognize embedded commands within the files.
- Dolphin Attack [4] has been shown to compromise voice assistants by modulating voice commands into ultrasonic frequencies, rendering them in audible to humans.

[16]
[14]
[17]
[18]
[10]
[11]
[12]

Methodology

Survey: Online Questionnaire

We believe the personal nature of voice assistants means that users are less cautious when using them. Our survey aims to test these following null (H_0) and alternative hypothesis (H_1) statements:

$\boldsymbol{H_0}$: Users adopt the same security posture when using voice assistants as they do when engaging with other technology and online communications.

$\boldsymbol{H_1}$: Users adopt a weaker security posture when using voice assistants as they do when engaging with other technology and online communications.

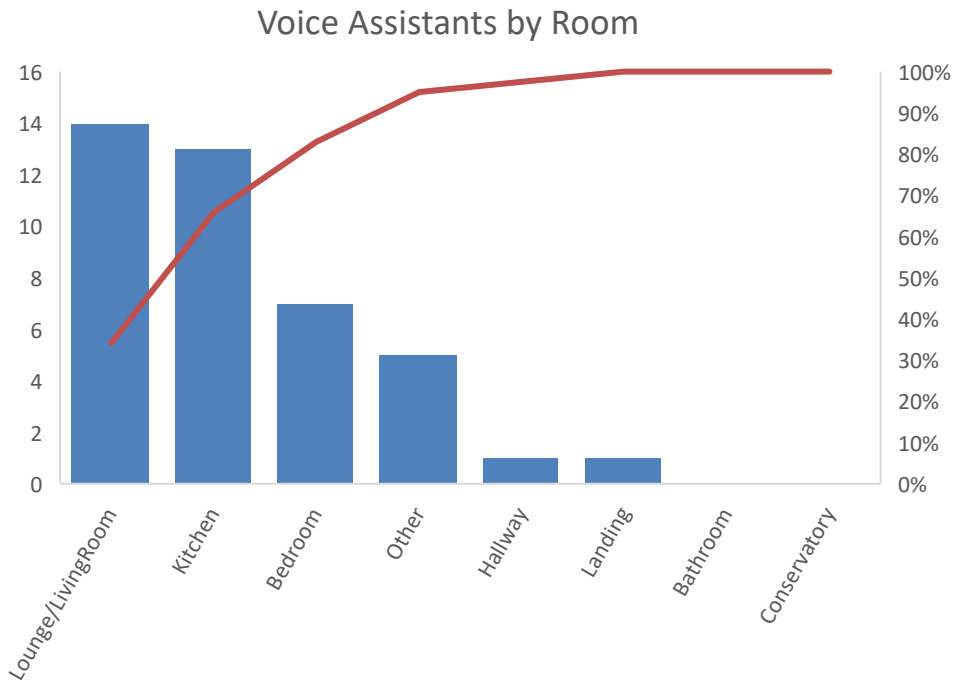
Our questionnaire gave focus to topics of behaviour, attitudes to security, and privacy. The full details of the questionnaire are available from our repository.

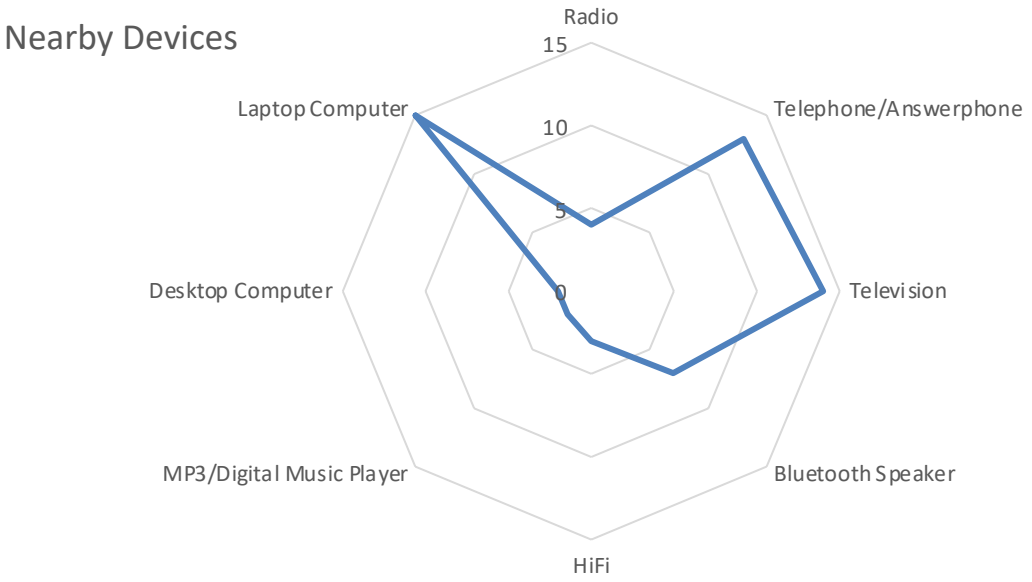
Experiments:

A selection of exploratory experiments in order to determine the seriousness of attacks and how easy they are; culminating in a proof-of-concept demonstration of how audio samples could be deliberately manipulated to trigger voice assistants, using adversarial noise.

- Shouting Through Letterboxes
- Replay Attack
- Answerphone
- TV, Laptop, other devices
- Camouflaged attacks
- Adding Adversarial Noise

Survey: Results





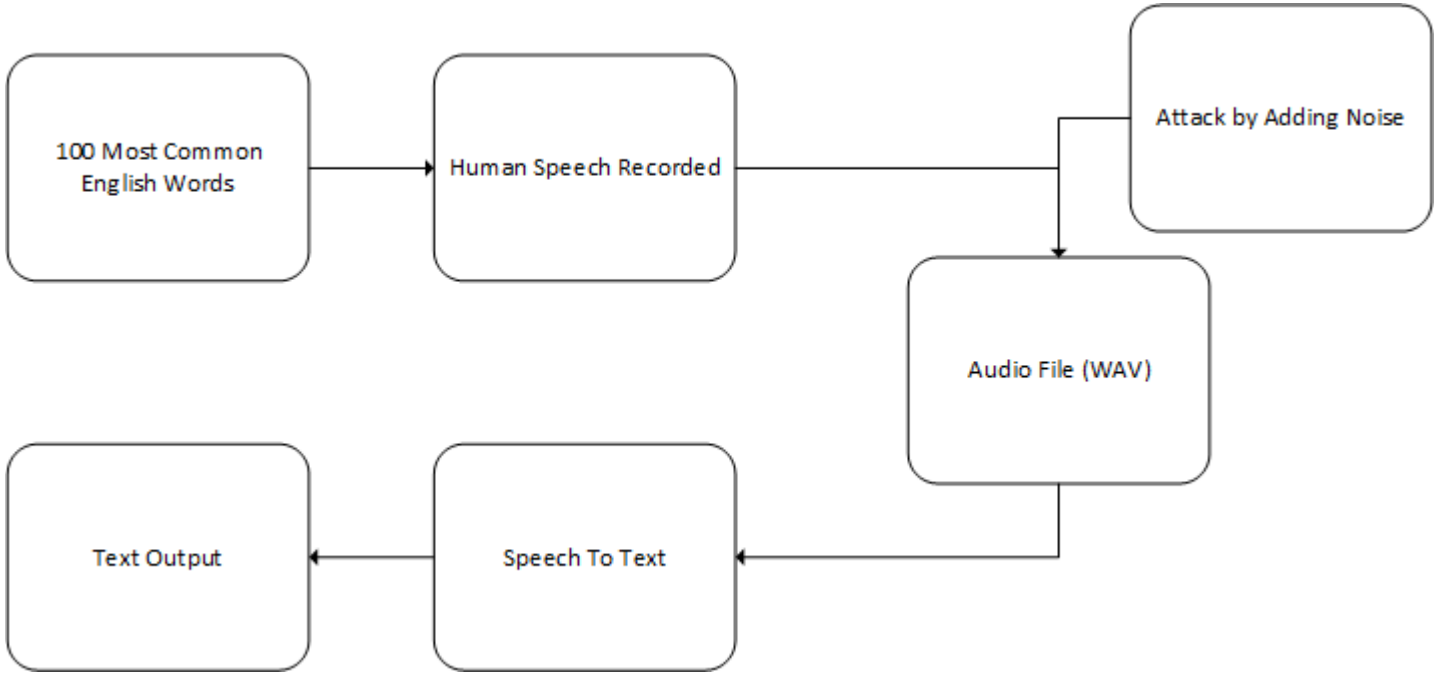
Shouting Through Letterboxes

Experiment	Result
Shouting Through Letterboxes	Triggers
Replay Attack	Triggers
Answerphone	Triggers
TV/Other Device	Triggers
Camouflaged	Triggers

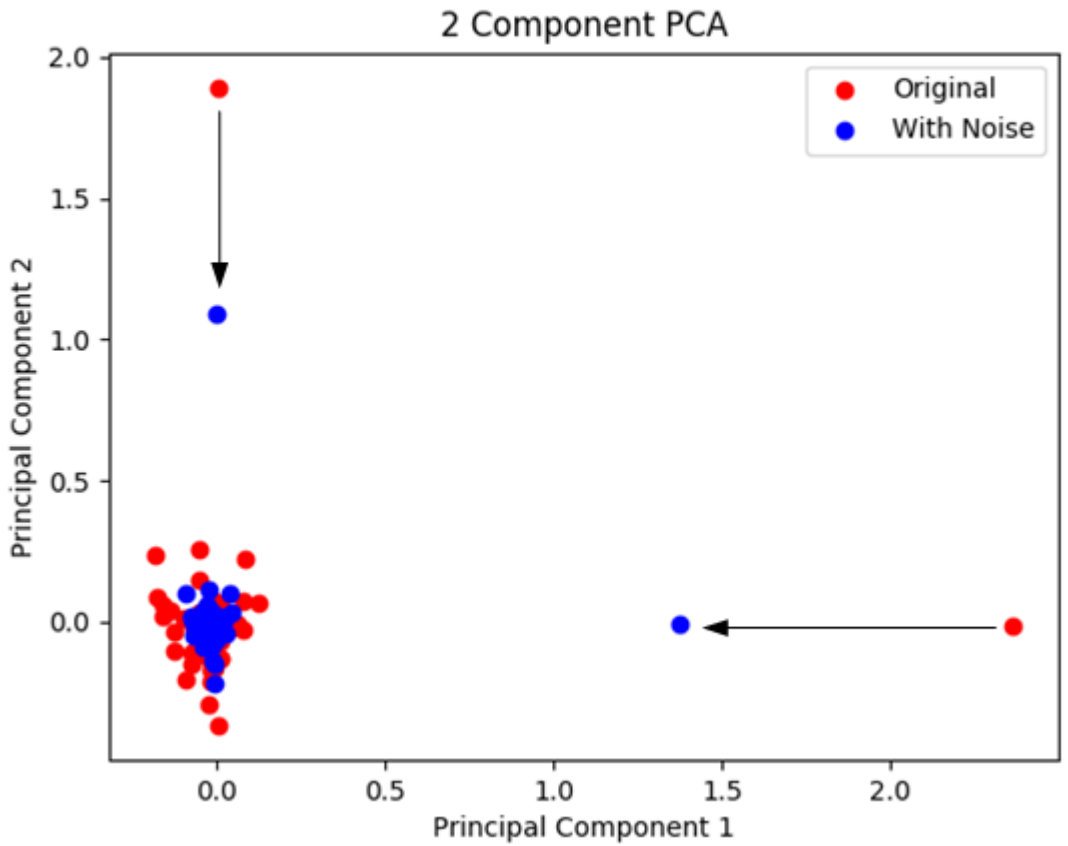
Adding Adversarial Noise

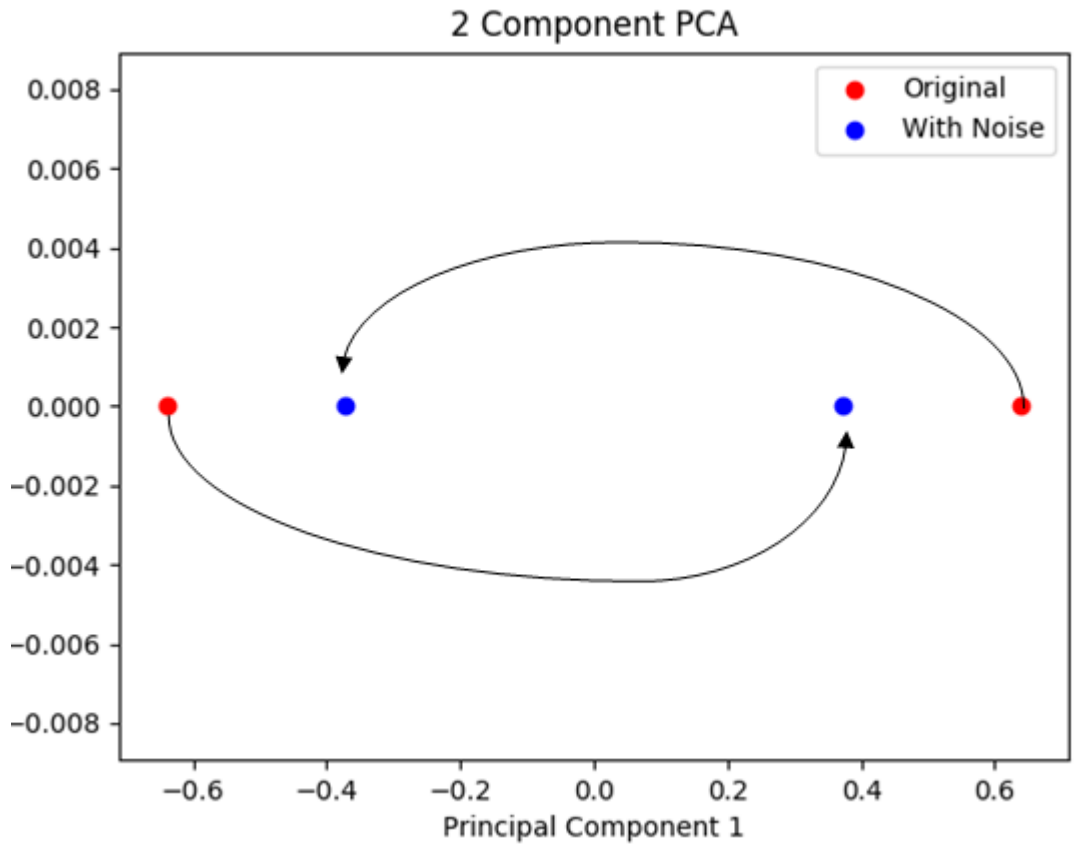
- 100 Common English Words
- Recorded audio files recorded on my laptop
- Python script using the SpeechRecognition API
- Subsequently dense white noise added
- Comparison of original and noisy samples showed google speech recognition system can be influenced through adversarial noise.

- It is difficult to assess the overall variability between samples
- Principal Component Analysis was used to perform dimensionality reduction, providing a more intuitive approach to compare samples.



Principal Component Analysis – Think & Thing



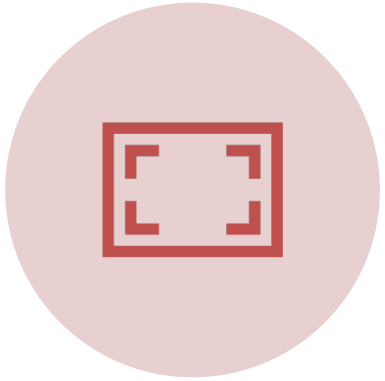


a) Principal Component analysis of 200 audio files.

b) Principal Component analysis of four audio files relating to the words Think and Thing.

Limitations

SAMPLING WAS NON-RANDOM



SIGNIFICANT MALE GENDER BIAS

THE SAMPLE SIZE IS RELATIVELY SMALL

Conclusions

- We show that a system could manipulate an audio sample, such that the difference is inaudible to but causes the output of a machine learning model to differ significantly.
- Our survey shows that participants do not mute devices and do not train their assistants to uniquely recognize their voice
- H_1 : Users adopt a weaker security posture when using voice assistants as they do when engaging with other technology and online communications

Takeaways

Voice as a user interface is inherently insecure.

The threat to voice assistants is an underrecognized risk.

Practice is at odds with stated preferences.

Adversarial noise attacks are viable. It is possible to nudge audio samples closer to a decision boundary.

Voice Assistants can be triggered by nearby devices.

Future Work

- Our future work investigates the nature of adversarial attacks in machine learning, and how greater protections can be developed combating threats, through identifying false inputs before they are accepted as input to learning models.
- Questions?
- If you have any other questions or comments please e-mail: Andrew6.McCarthy@uwe.ac.uk

MORE EXPERIMENTATION

HOW TO MODIFY AN AUDIO SAMPLE SO THAT IT FULLY RECOGNIZED AS A VOICE COMMAND.

https://mccarthya-s3-bucket.s3.eu-west-2.amazonaws.com/publications/ShoutingThroughLetterboxes/index.html
Twitter: @AndyMcCarthyUK

NO MORE THAN 12 SLIDES!

Questions?

https://mccarthya-s3-bucket.s3.eu-west-2.amazonaws.com/publications/ShoutingThroughLetterboxes/index.html

Twitter: @AndyMcCarthyUK

Thanks and Acknowledgements

Fellow Authors:

Phil Legg and Benedict Gaster

Supervisory Team

Dr. Phil Legg, Associate Professor in Cyber Security, UWE; Dr. Panos Andriotis, Senior Lecturer in Computer Forensics and Security, UWE; Dr. Essam Ghadafi, Senior Lecturer in Computer Science, UWE; Dr. Larry Bull, Professor of Artificial Intelligence, UWE.

Acknowledgements

I would like to thank Techmodal Ltd for supporting this research.

Motivation

- Household names: Alexa, Google, Cortana
- Smart Speakers
- Phones & Tablets
- Traditional Desktops and Laptops

Many of us really enjoy the convenience of using such assistants; we have focussed on the benefits of these systems without fully considering their vulnerabilities.

Motivation

- Household names: Alexa, Google, Cortana
- Smart Speakers
- Phones & Tablets
- Traditional Desktops and Laptops

Many of us really enjoy the convenience of using such assistants; we have focussed on the benefits of these systems without fully considering their vulnerabilities.

Why this research is important

- Household names: Alexa, Google, Cortana
- Smart Speakers
- Phones & Tablets
- Traditional Desktops and Laptops

Many of us really enjoy the convenience of using such assistants; we have focussed on the benefits of these systems without fully considering their vulnerabilities.

Ubiquity of Voice Assistants

- Household names: Alexa, Google, Cortana
- Smart Speakers
- Phones & Tablets
- Traditional Desktops and Laptops

Many of us really enjoy the convenience of using such assistants; we have focussed on the benefits of these systems without fully considering their vulnerabilities.