Building an AI meetings assistant

Anna Cowan
MSc in Big Data

Problem

Minute taking and the recording of actions and decisions in meetings can be time consuming and distracting. This project provides an approach for recording decisions and actions from speech by using designated keywords. Using the starter keyword ("Hey Sharkey") and end point keyword ("Bye Sharkey") will likely help participants make sure their points are clear and well formulated before continuing.

Aim

To extract the decisions and actions from any meeting audio where the specific keywords are used.

5

Example Output

Decisions made:

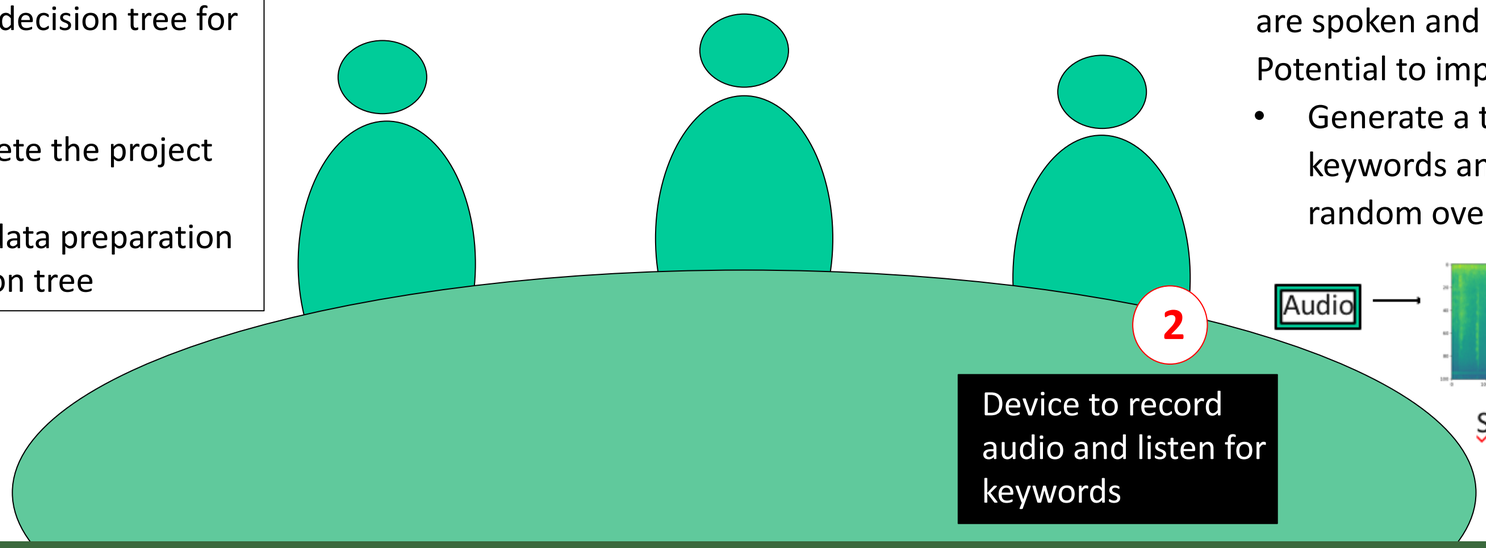
- We will use a decision tree for classification

Action Items:

- Anna - Complete the project poster
- Joe – Do the data preparation for the decision tree

…
A: "Thanks that would be great Hey Sharkey Anna will complete the project poster Bye Sharkey"
C: "Let's discuss the approach for… "
…
B: "Hey Sharkey We will use a decision tree for classification Bye Sharkey. Joe would you mind preparing the data"
C: "Yes I can do that"
B: "Hey Sharkey Joe will do the data preparation for the decision tree Bye Sharkey"
A: "Great I think that's all."

1



2

Device to record audio and listen for keywords

3

Approach

- Using a speech-to-text API to gather the text from audio, and use the transcript to search for keywords.
- Exploring custom language models to increase the ability to recognise keywords.
- Using NLP techniques to determine if extracted text was a decision or action Classification problem, using:
  - Sentence parsing
  - Part – of – speech tagging
  - Named entity recognition
  - tf–idf

Example challenge:

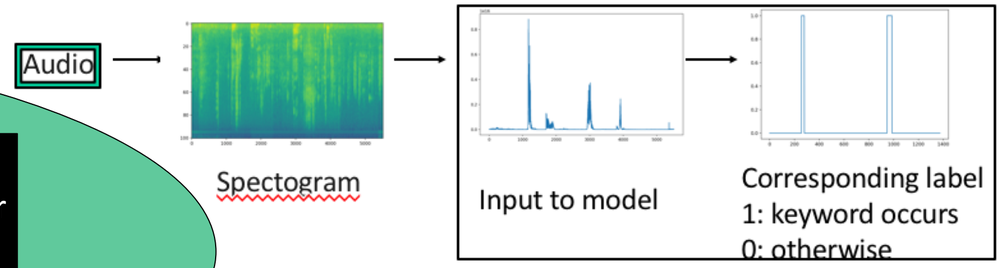
File: t_0.wav Text: **the shocking** Confidence: 0.265
File: t_1.wav Text: **Hey Sharkey** Confidence: 0.08
File: t_2.wav Text: **Hey Shockey** Confidence: 0.417
File: t_3.wav Text: **exactly** Confidence: 0.016
File: t_4.wav Text: **Hey Shockey** Confidence: 0.38
File: t_5.wav Text: **HRK** Confidence: 0.111
File: t_6.wav Text: **he sure K.** Confidence: 0.595
File: t_7.wav Text: **HRK** Confidence: 0.129
File: t_8.wav Text: **yeah sure K.** Confidence: 0.747
File: t_9.wav Text: **yeah sure K.** Confidence: 0.49
File: t_10.wav Text: **Hey Jackie** Confidence: 0.164
File: t_11.wav Text: **he checked the** Confidence: 0.561
File: t_12.wav Text: **Hey Jackie** Confidence: 0.456
File: t_13.wav Text: **he Sharkey** Confidence: 0.233
File: t_14.wav Text: **Hey Jackie** Confidence: 0.956
File: t_16.wav Text: **a shocking** Confidence: 0.696
File: t_17.wav Text: **it's shocking** Confidence: 0.73
File: t_18.wav Text: **a shocking** Confidence: 0.304
File: t_19.wav Text: **a shock** Confidence: 0.425
File: t_20.wav Text: **a shocking** Confidence: 0.332

IBM Watson transcriptions of "Hey Sharkey" audio

4

Possible Further Steps

- Build a recurrent neural network to detect when keywords are spoken and provide time stamps for where to transcribe. Potential to improve keyword detection.
  - Generate a training dataset by recording samples of the keywords and non-keyword words and combining at random over samples of background audio.



stir.ac.uk For further information contact: *Anna Cowan* *amc00076@students.stir.ac.uk*

BE THE DIFFERENCE