

Application of Keyword Spotting System on accurately predict memory test result

Author: Yanning Cao

Supervisor: Dr. Beena Ahmed

1. Problem statement:

- Memory test: Can help doctors to detect dementia such as Alzheimer's in early stages which is crucial in treatment.
- Keyword Spotting System: Detecting a specific word of interest from a continuous stream of audio.
- Aim of the research: Find out whether keyword spotting system is able to predict memory test results and analyze the performance of the KWS model.

3. Architecture of ResNet:

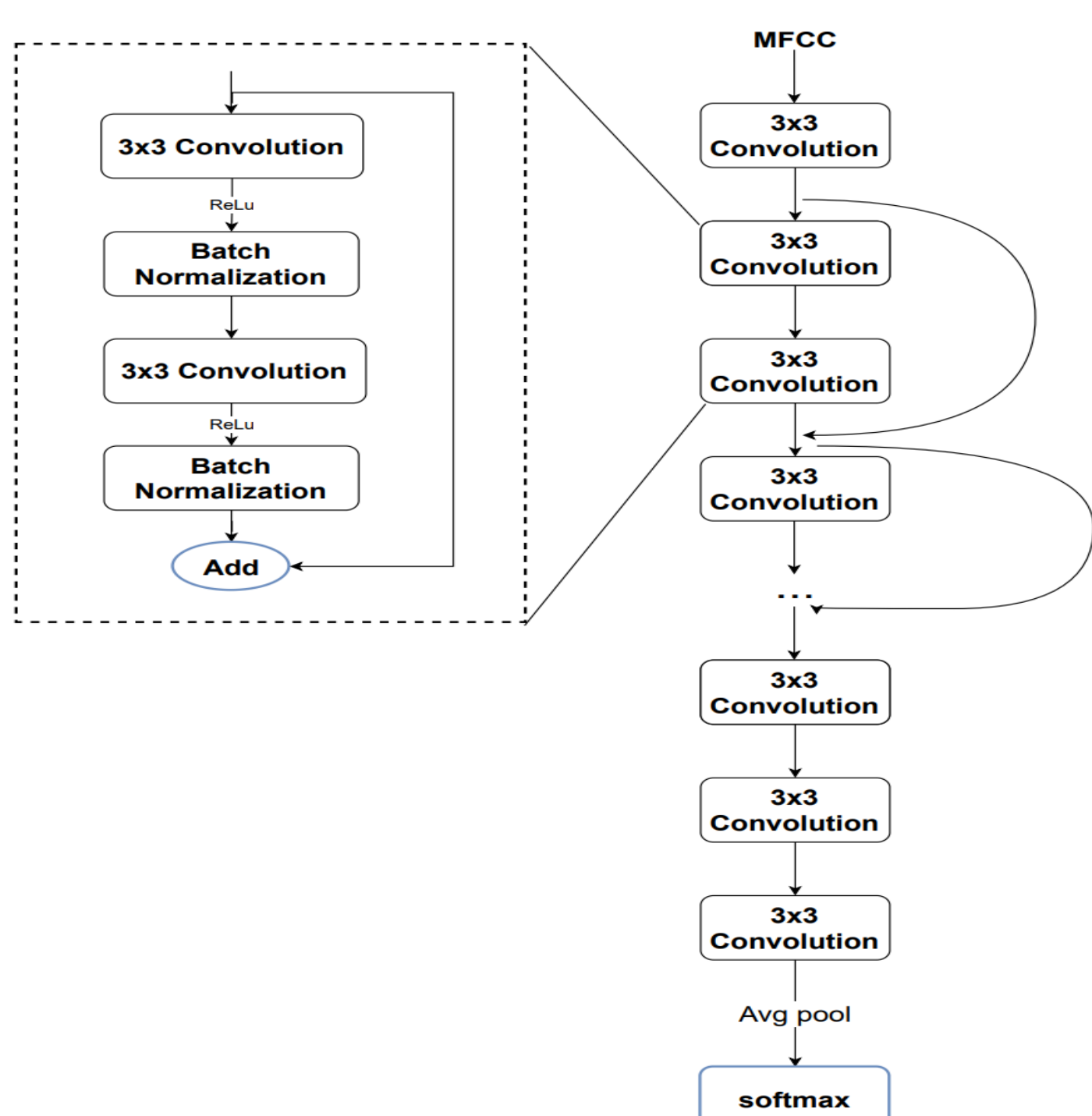


Fig 3. Architecture of ResNet Neural network framework

4. Datasets:

- Mozilla Common Voice**
 - 13,905 recorded hours of mp3 audio data
 - 11,192 validated hours in 76 languages
- Google Speech Commands Dataset**
 - 65,000 one-second-long utterances of 30 short words by thousands of different people
- Memory test result dataset**
 - NHMRC funded Maintain Your Brain (MYB) project
 - 14296 telephone audio recordings across 4 trials from 4085 participants aged between 55 and 77

2. Overview of Keyword Spotting System:

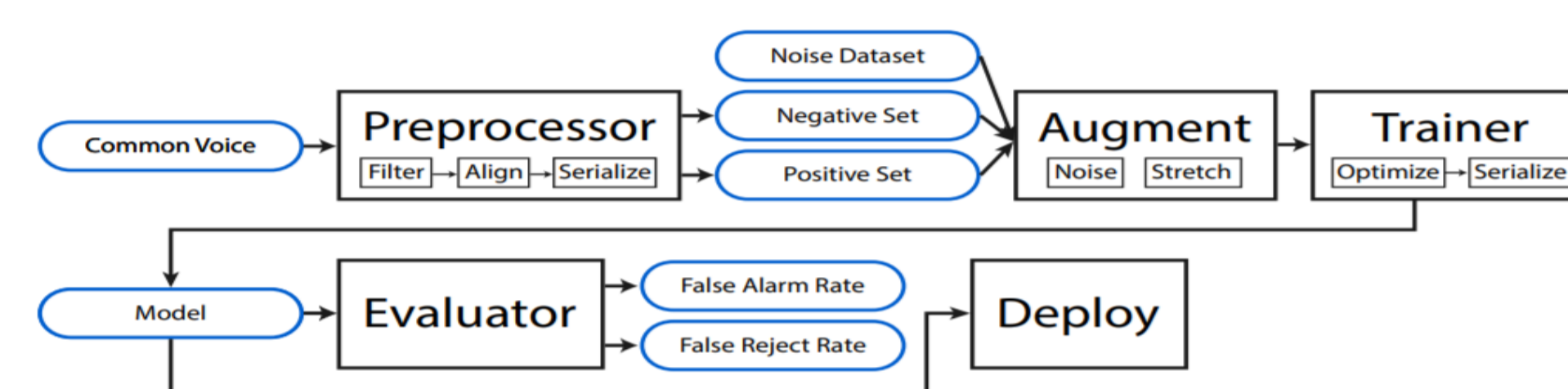


Fig 1. An illustration of Keyword Spotting system's framework and its control flow.

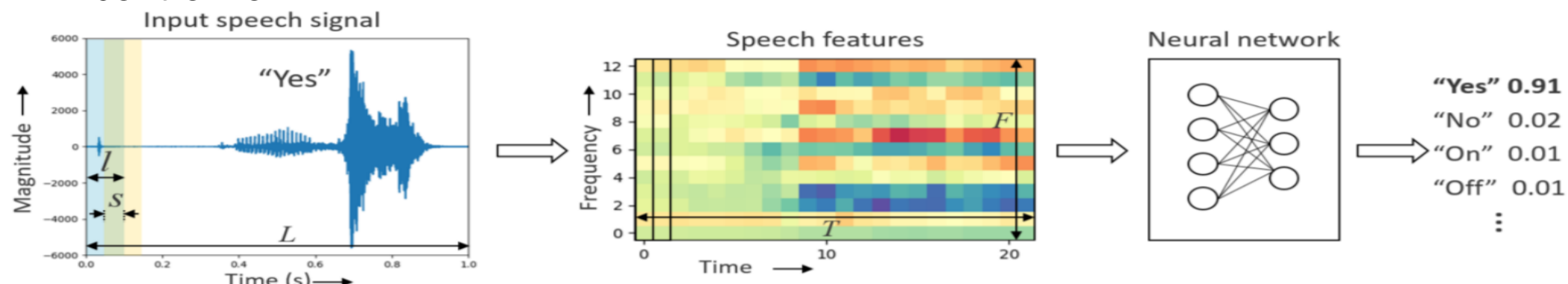
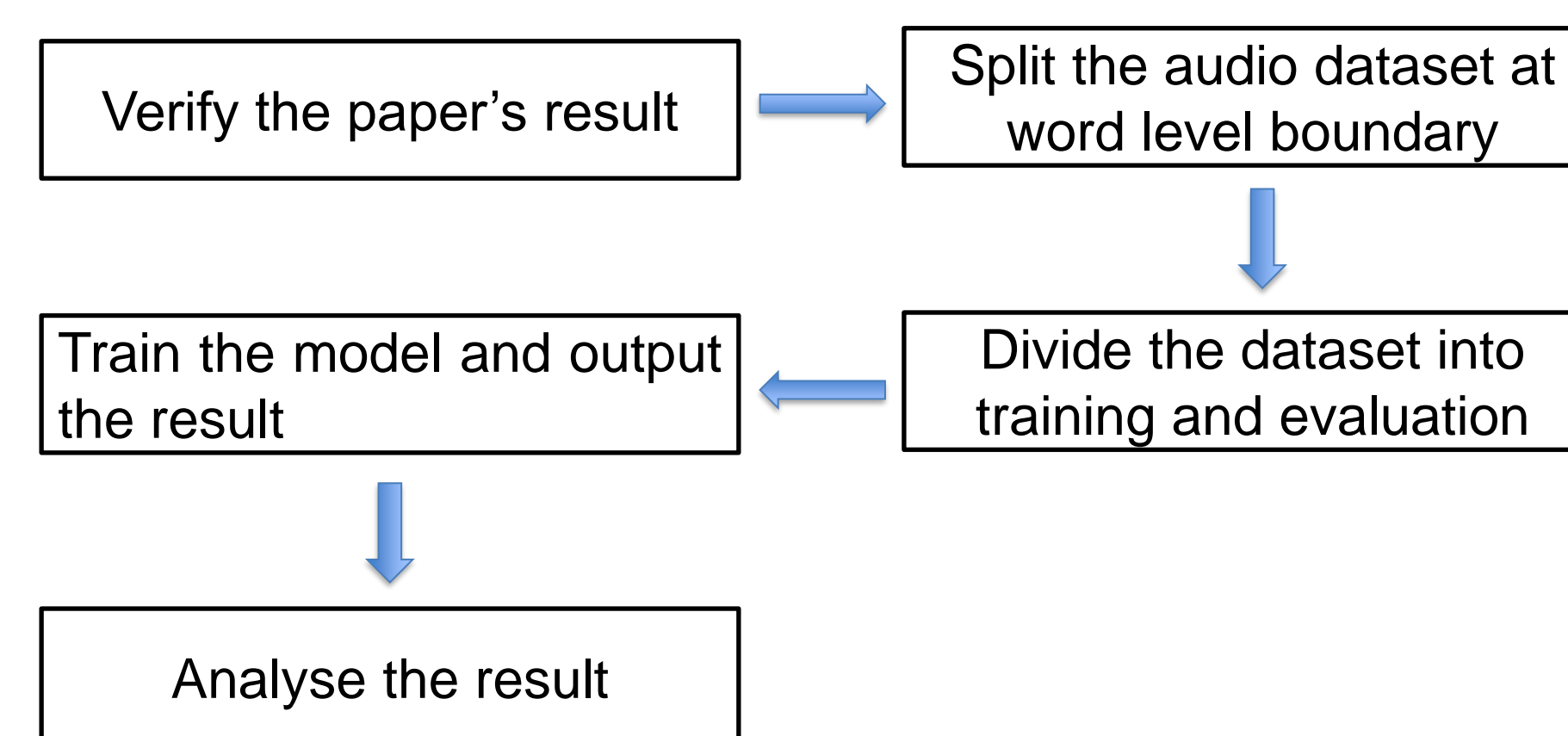


Fig 2. Keyword Spotting system's illustration

5. Experimental process:



Experiment 1: Verification of Paper's result and data preparation

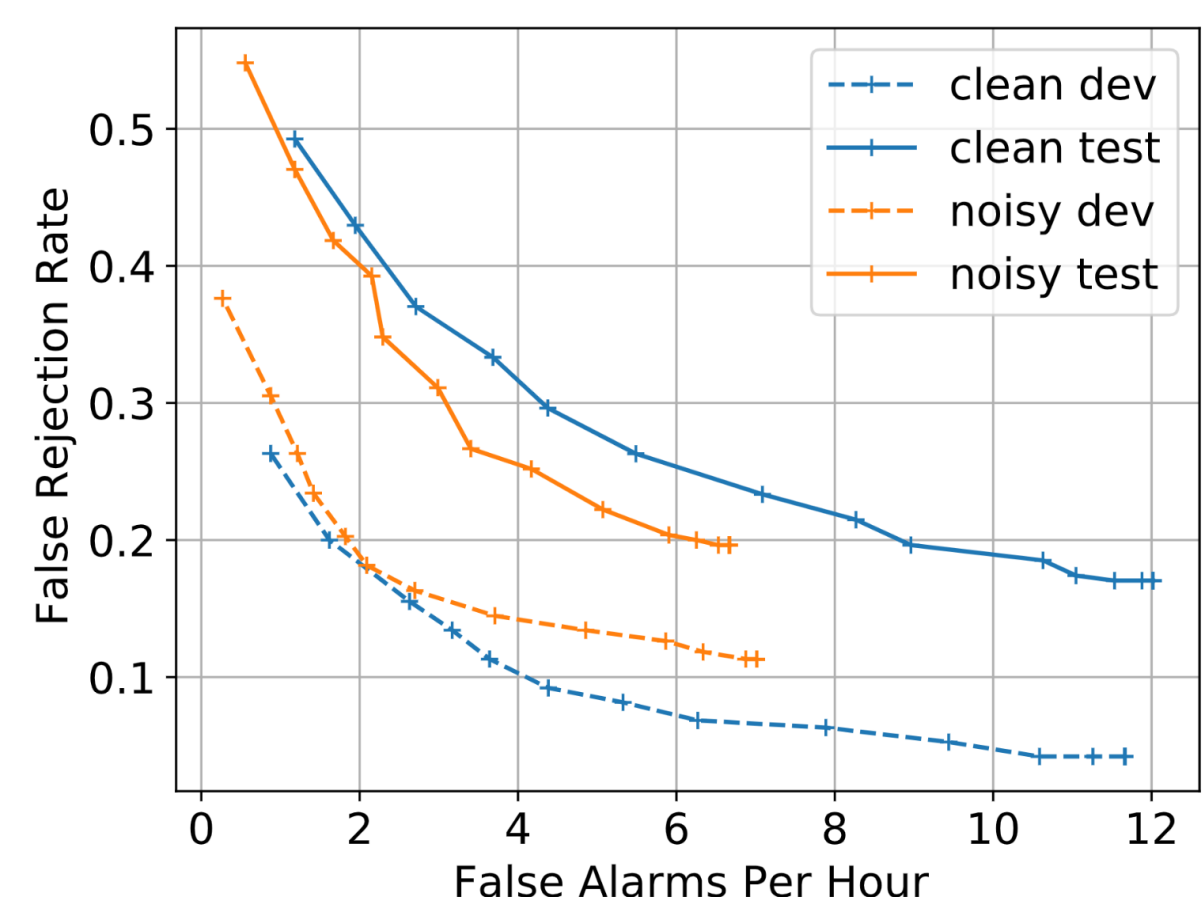


Fig 4. Verify thesis result on Firefox dataset

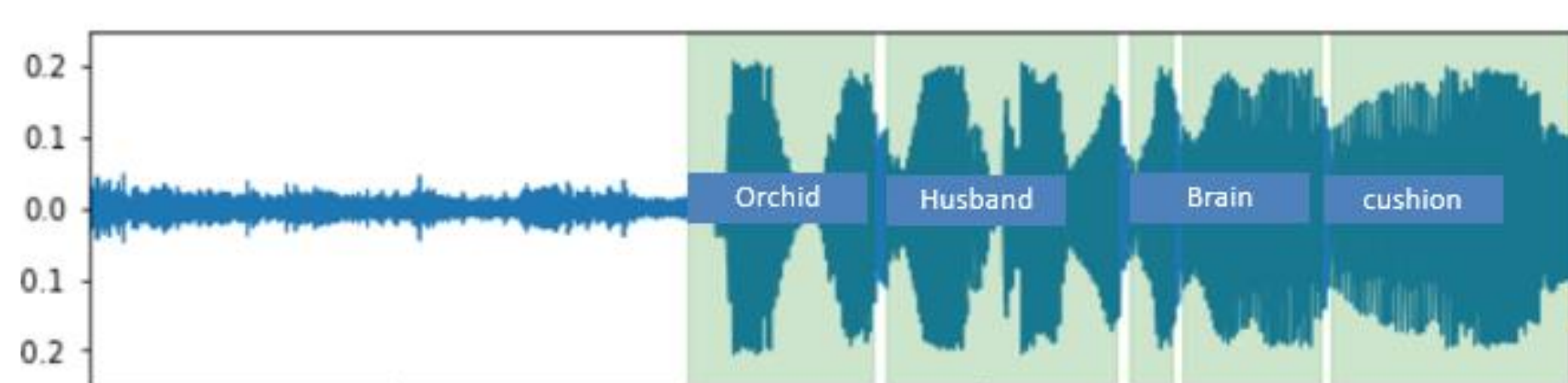


Fig 5. Split the audio dataset at word level boundary

	Mozilla Common Voice	Google command dataset	Memory test result dataset
Accuracy	96%	97%	87%

Fig 5. KWS average performance on different dataset

Experiment 2: Why using Keyword Spotting system

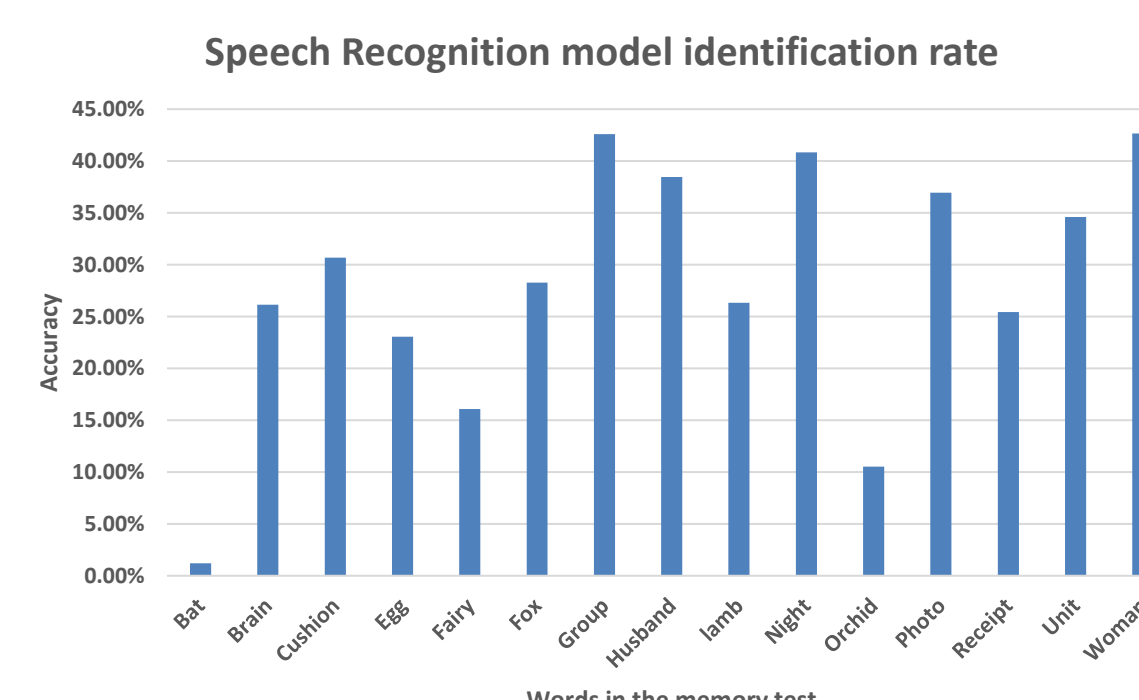


Fig 6. Performance using the speech recognition model

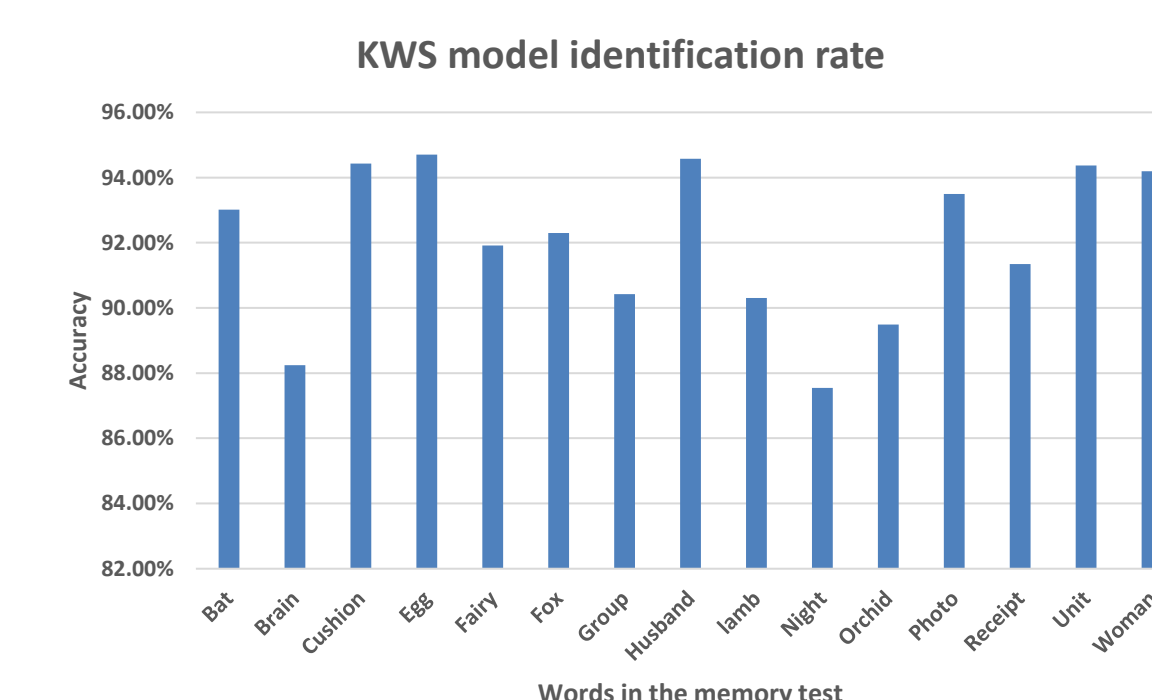


Fig 7. Performance using the Keyword Spotting System

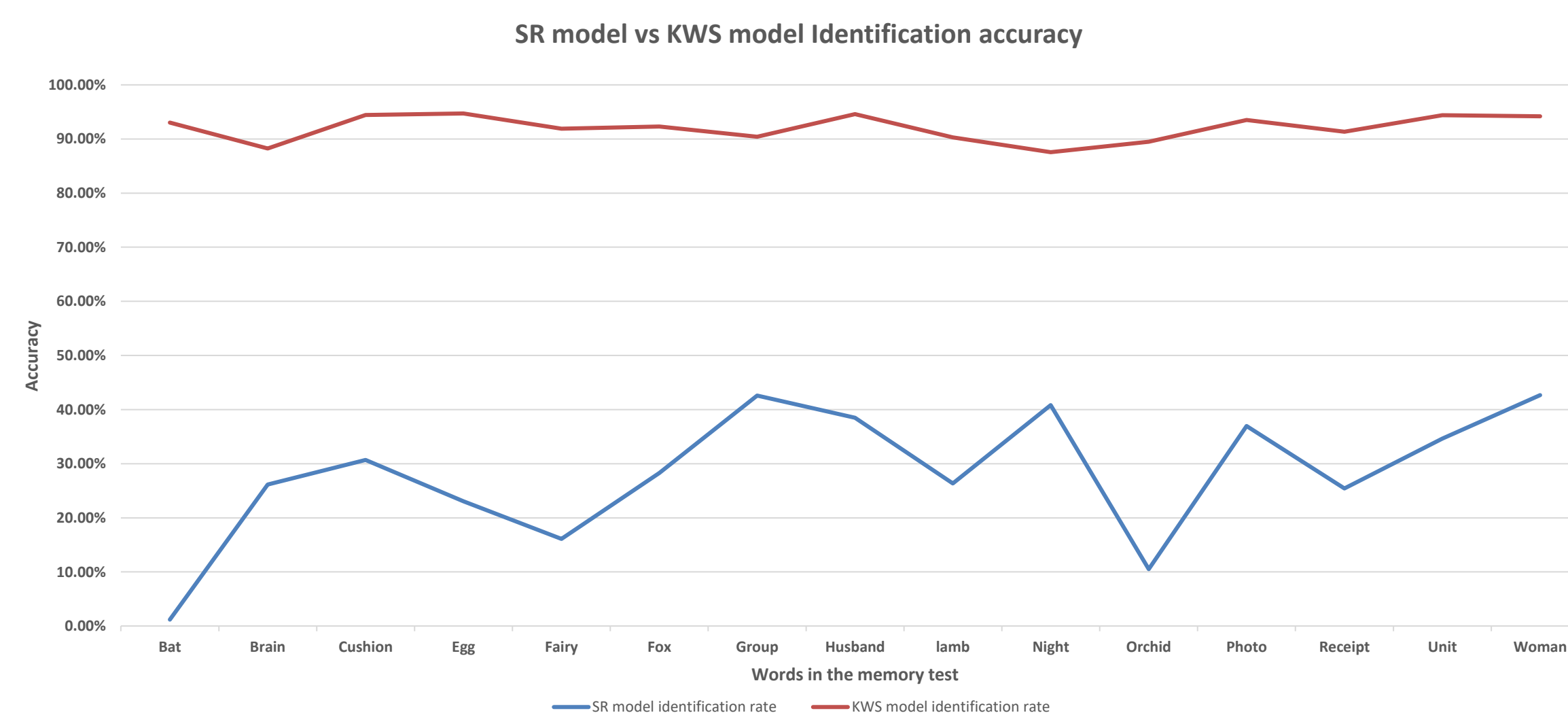


Fig 8. Performance difference between Speech recognition model and Keyword Spotting model

Experiment 3: Impact of dataset size and Split ratio on system's performance

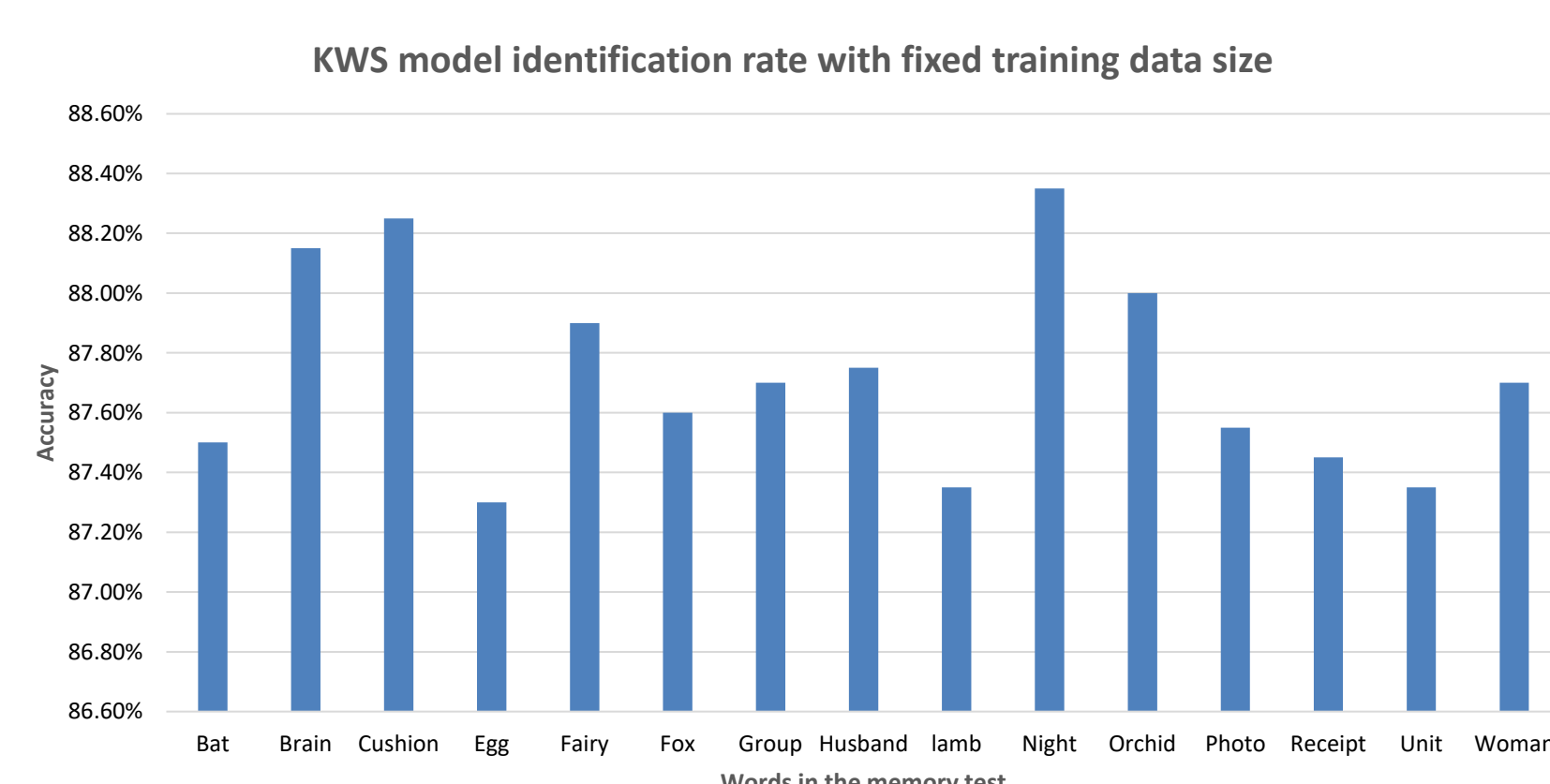


Fig 9. Accuracy on each word using 2000 sample as training size

- Definitions:
- Training dataset
 - Validation dataset
 - Split ratio = $\frac{\text{training dataset size}}{\text{validation dataset size}}$

- Experimental Process:
- Control variates with:
 - Words
 - Dataset size
 - Split ratio

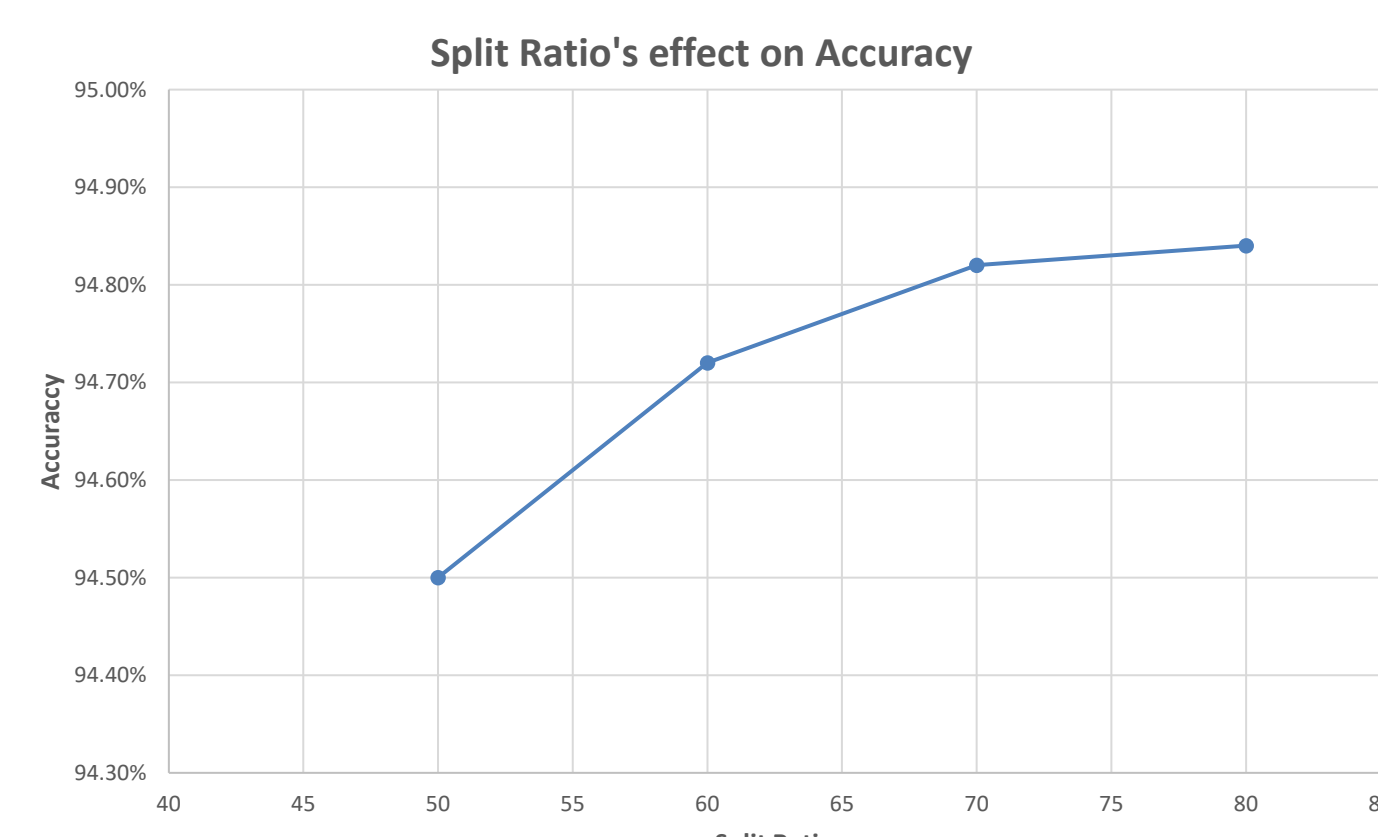


Fig 10. Different Split ratio's impact on performance

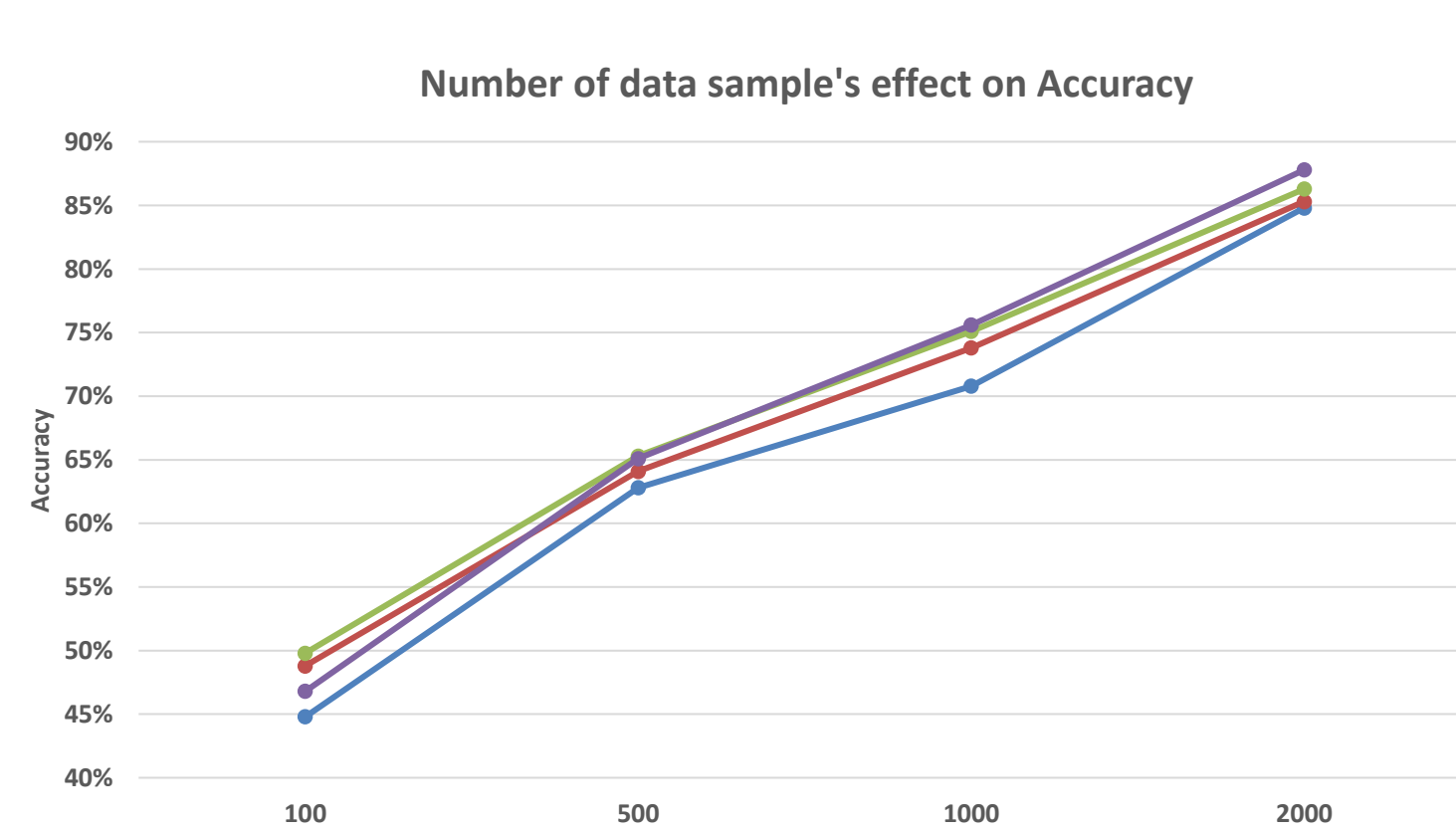


Fig 11. Different dataset size's impact performance

Conclusion:

- Keyword Spotting System perform decently on predicting the memory test result
- Keyword Spotting System have better performance compares to the Speech recognition system
- All of the words in the memory test result have similar accuracy.
- Both number of samples and split ratio have impact on the performance of the model

Future works:

- Integrate the dataset generation process into the system
- Increase noise robustness of the system
- Further classify the dataset (gender age) and study the result