Autonomous Vehicle Speaker Verification System (AVSVS)

By: Aaron Pfalzgraf and Chris Sullivan
Project Advisor: Dr. Jose Sanchez

Presentation Outline

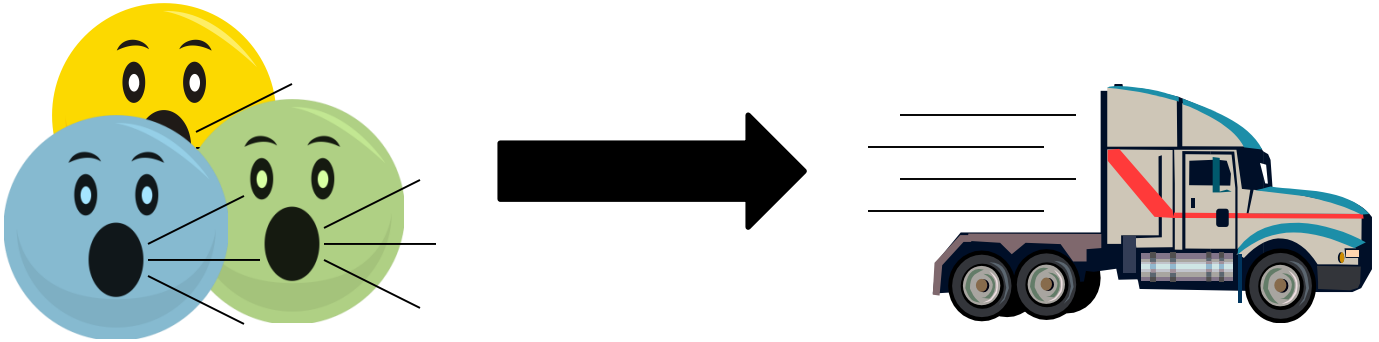
- Background
- Relevance and Motivation
- Project Overview and Specifications
- Methods
- Results
- Conclusion
- References

Speaker Verification Background

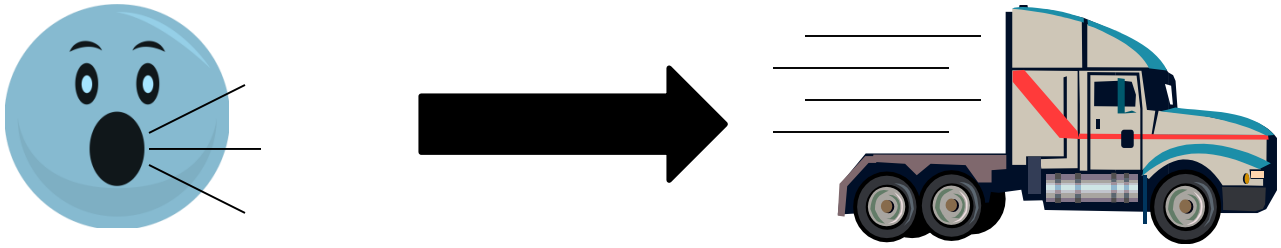
- Identify people by the sound of their voices
- Alternative to security passwords
- Similar system to speech recognition
- Text-dependent vs. text-independent

Relevance and Motivation

- Voice command system security threats:

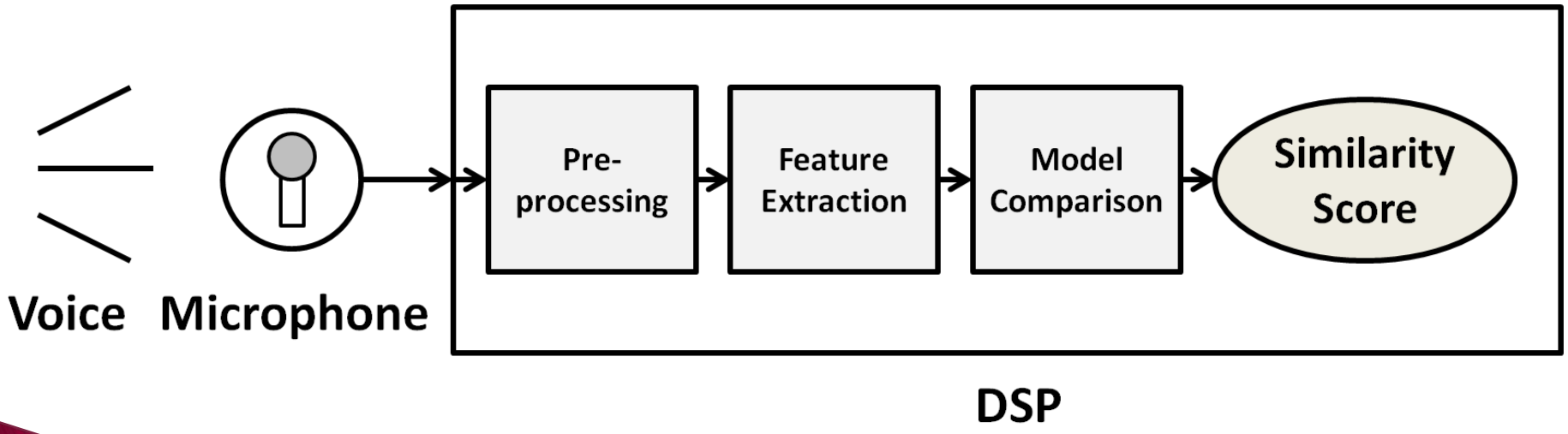


- Threats mitigated by speaker verification:

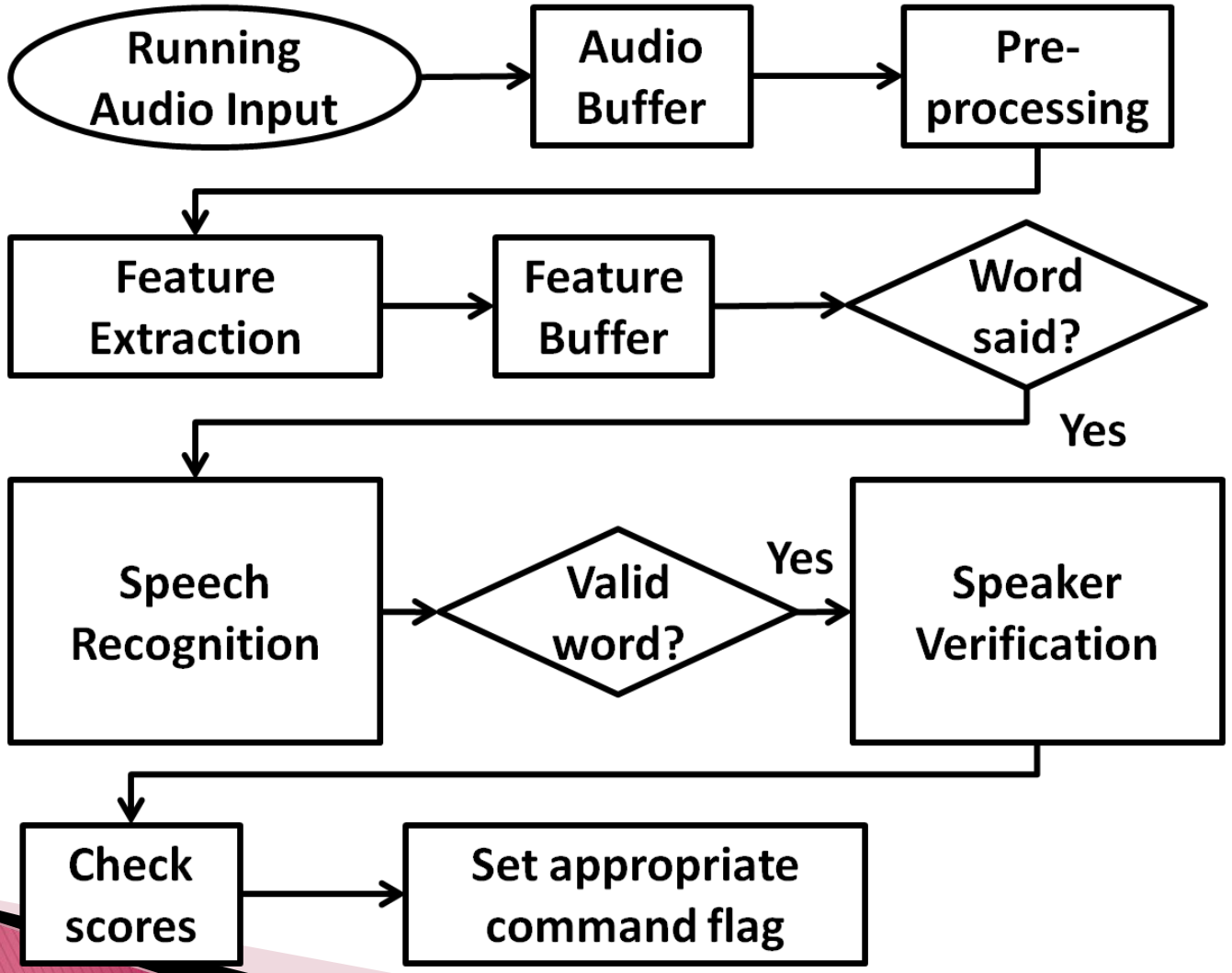


Project Overview

- Integrate speech recognition with speaker verification
- Operate in real time on a digital signal processor



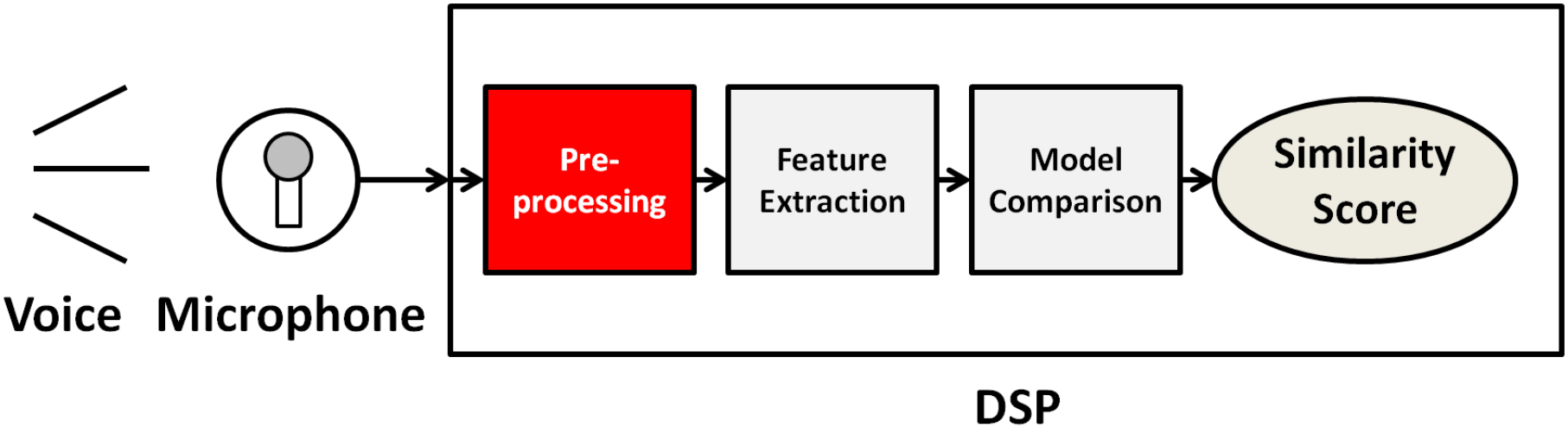
Project Overview



Project Specifications

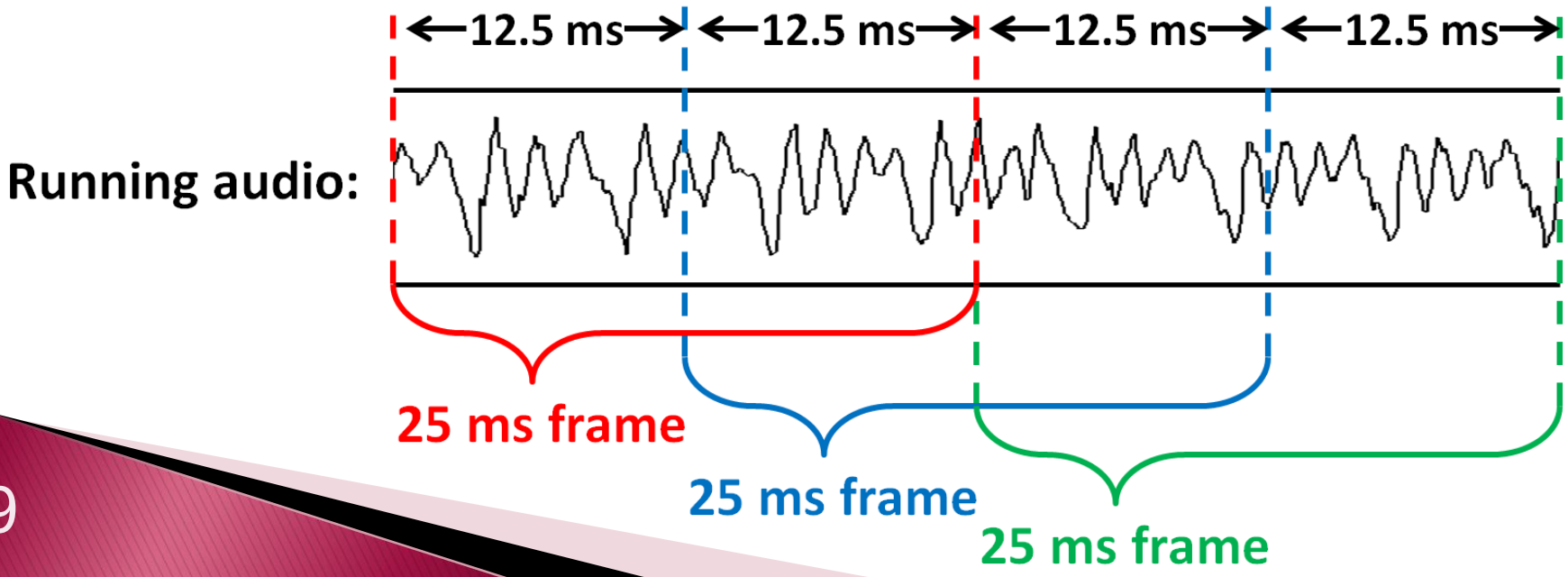
- True speaker rejection under 1%
- Imposter acceptance preferably under 2%
- 8 kHz audio sampling rate
- System response time under 50 ms
- Implemented on TI eZdsp5535 DSP

Pre-processing



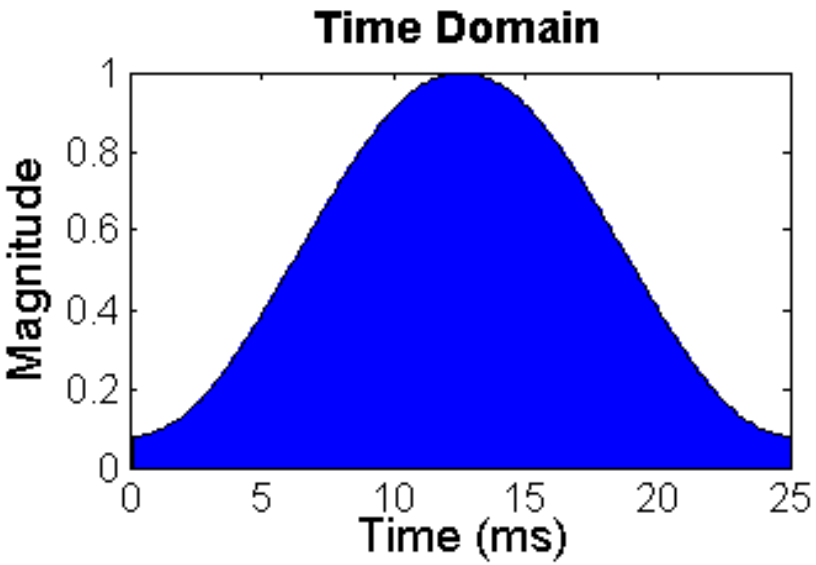
Pre-processing

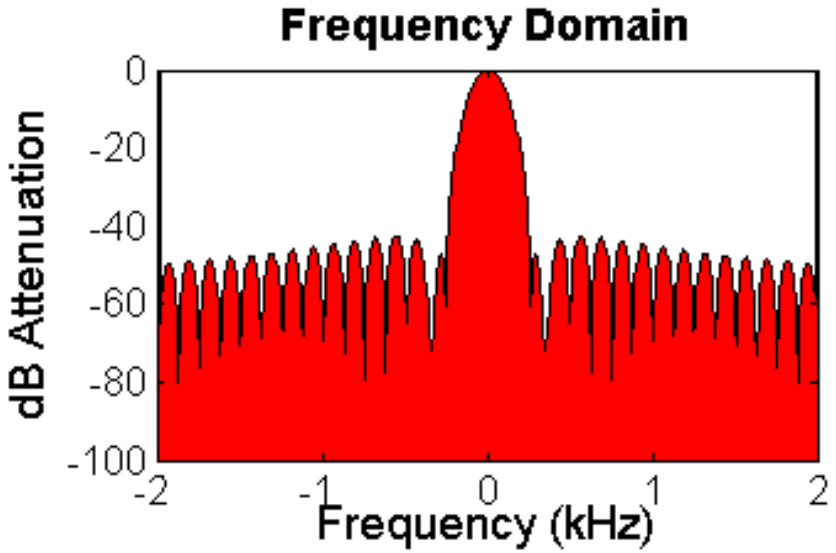
- Audio stored in 25 ms circular buffer
- 50% frame overlap
- Neglect frames of silence



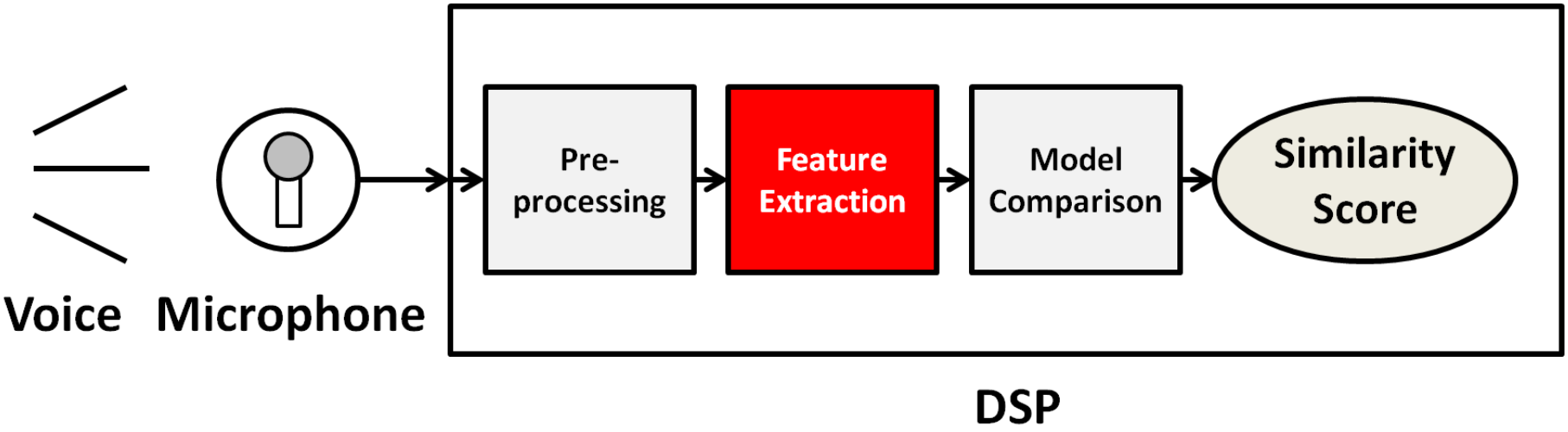
Pre-processing

- Hamming window characteristics:
  - First side-lobe attenuation: -43 dB
  - Main lobe width: 1.30 frequency bins





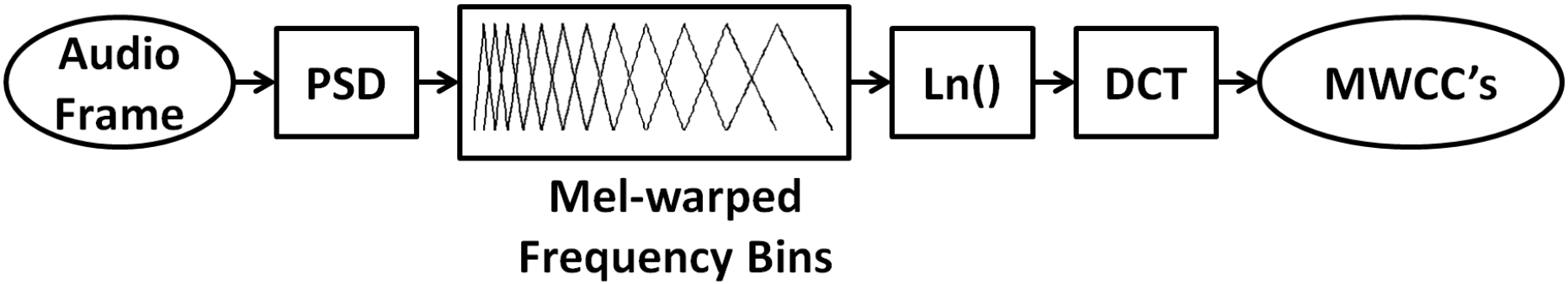
Feature Extraction



Feature Extraction

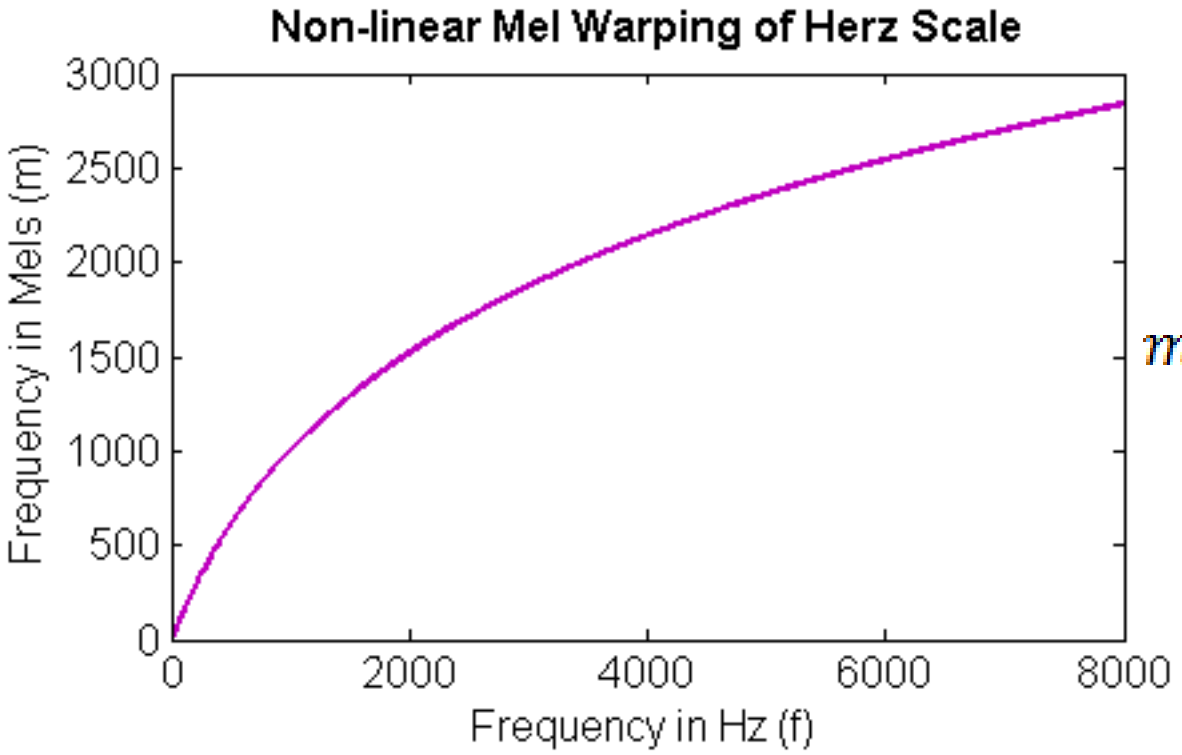
- 15 Mel-warped cepstral coefficients (MWCC) extracted per frame

- Measure of short term power spectral density



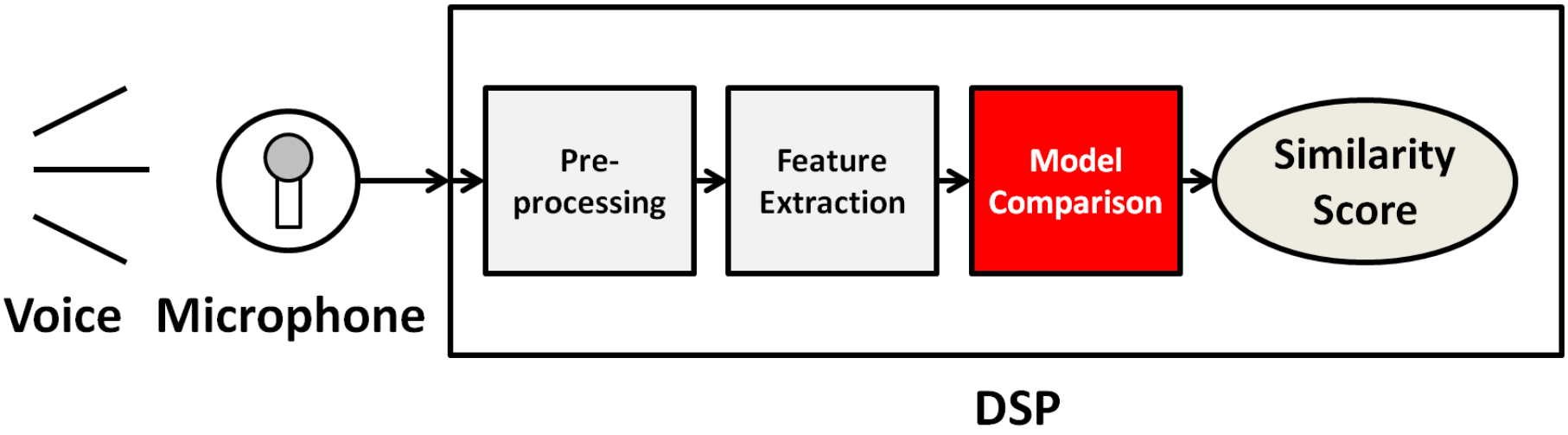
Feature Extraction

- Mel scale mimics response of the human ear



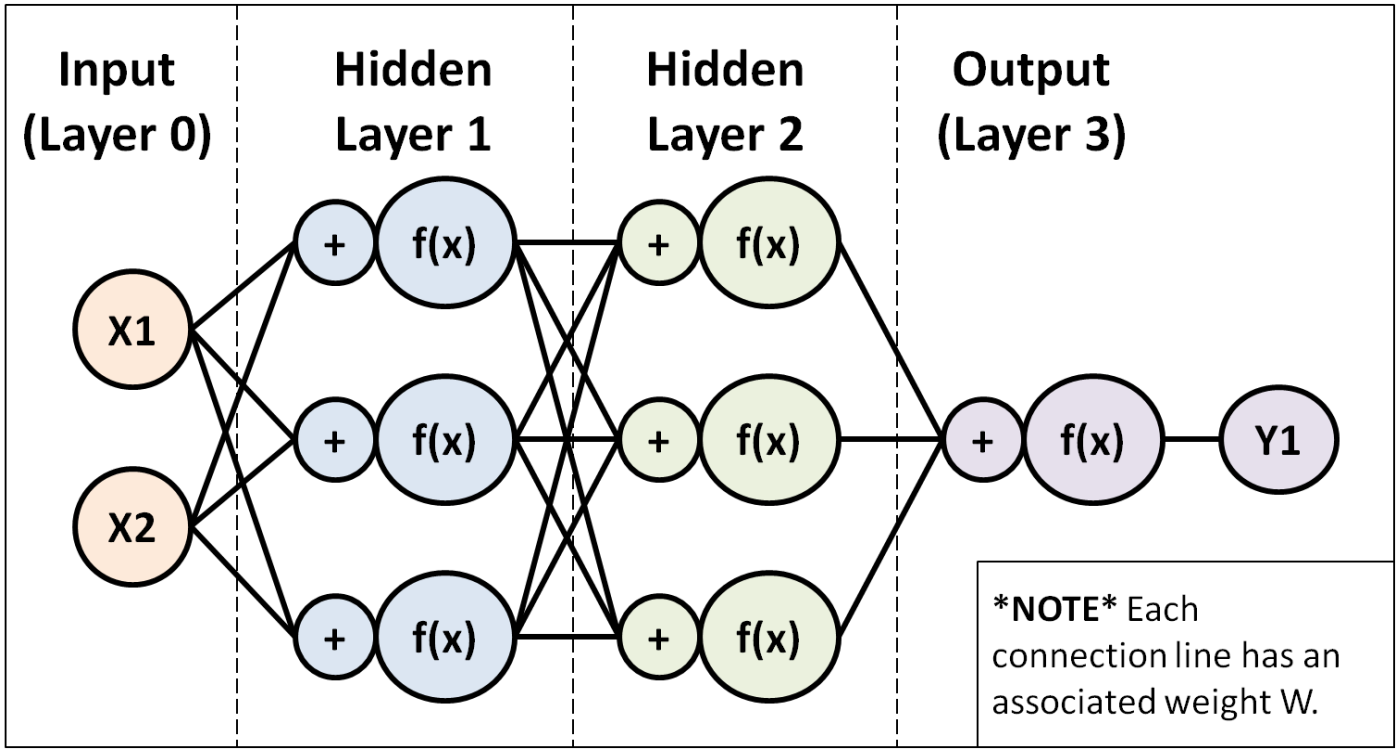
$$m = 2595 \log_{10}(1 + \frac{f}{700})$$

Model Comparison

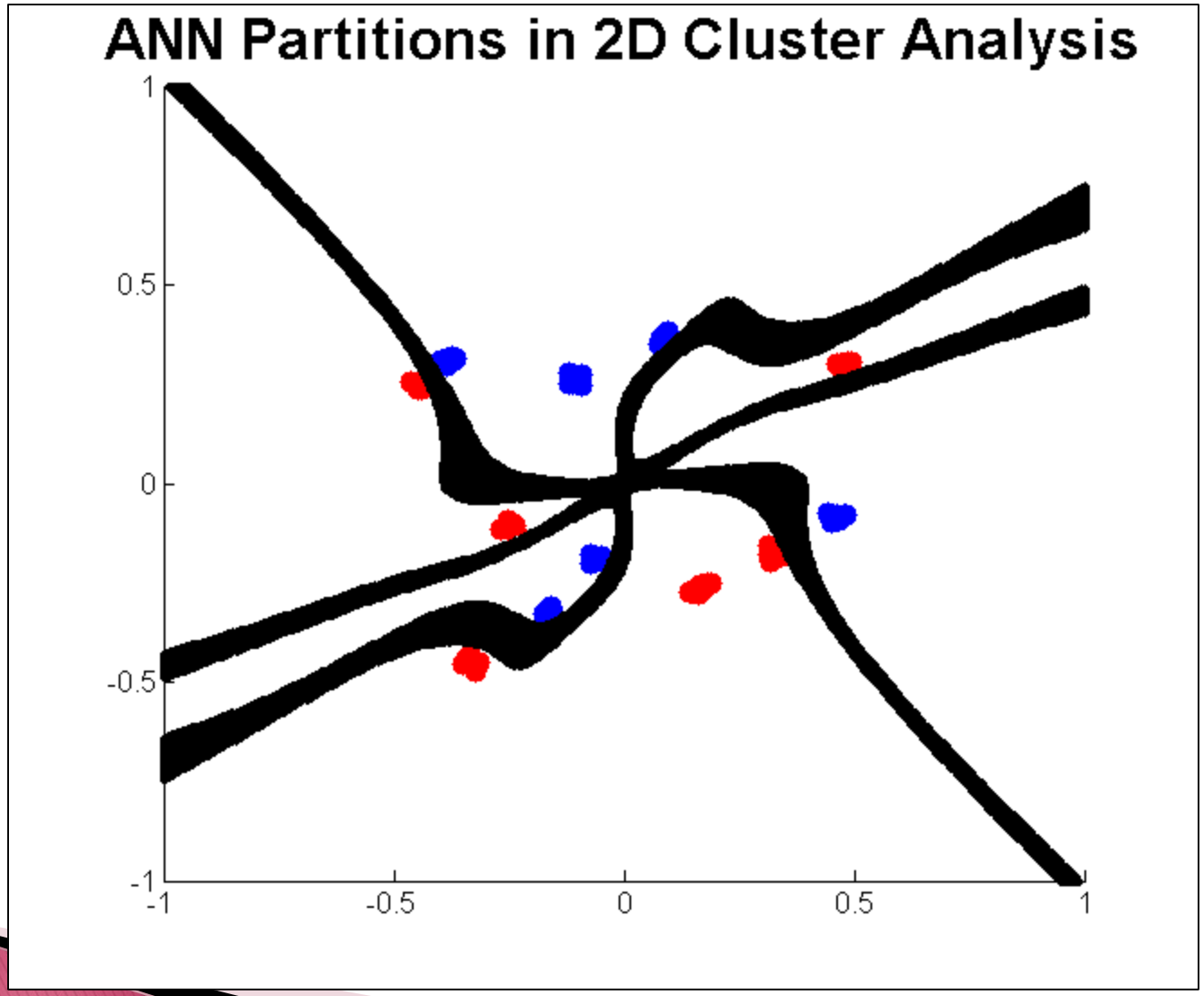


Model Comparison

- Artificial Neural Networks (ANN) perform cluster analysis

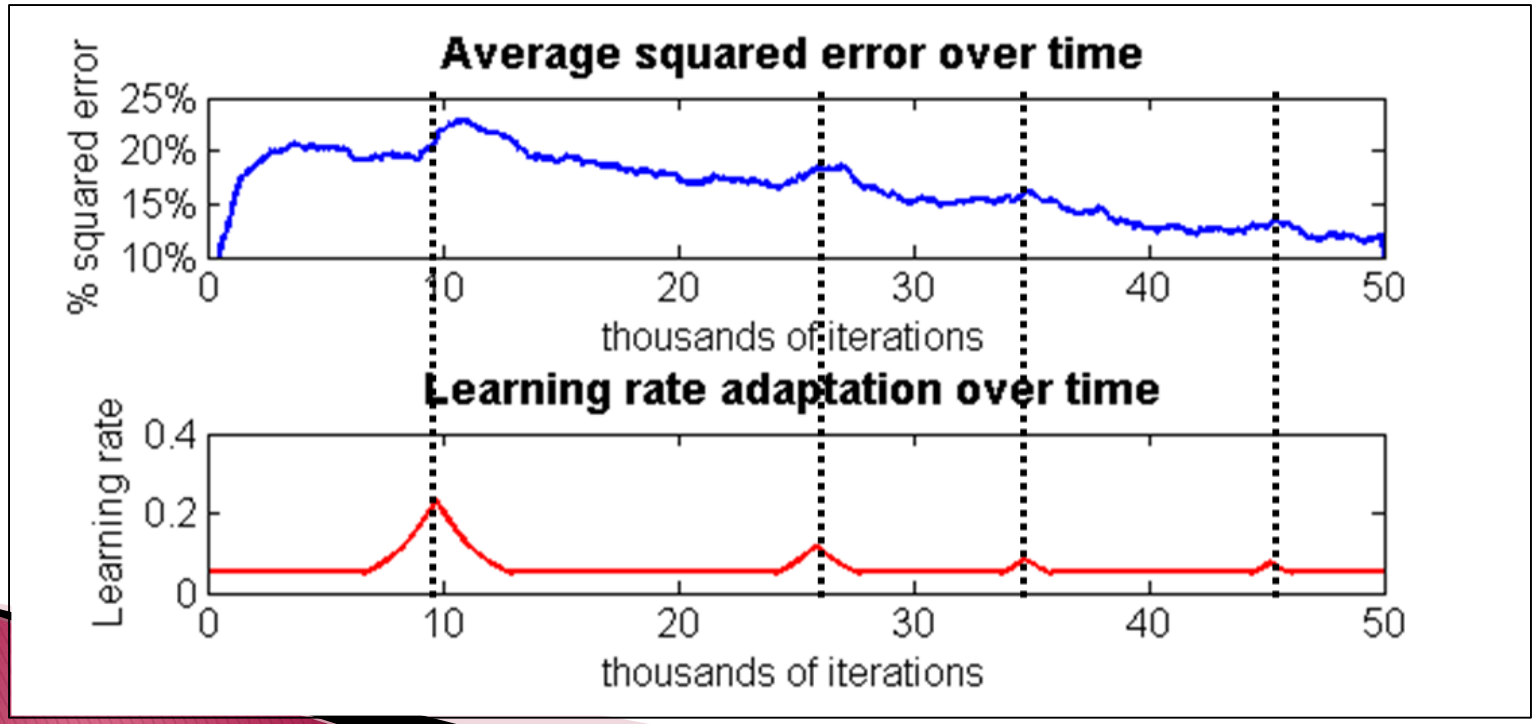


Model Comparison

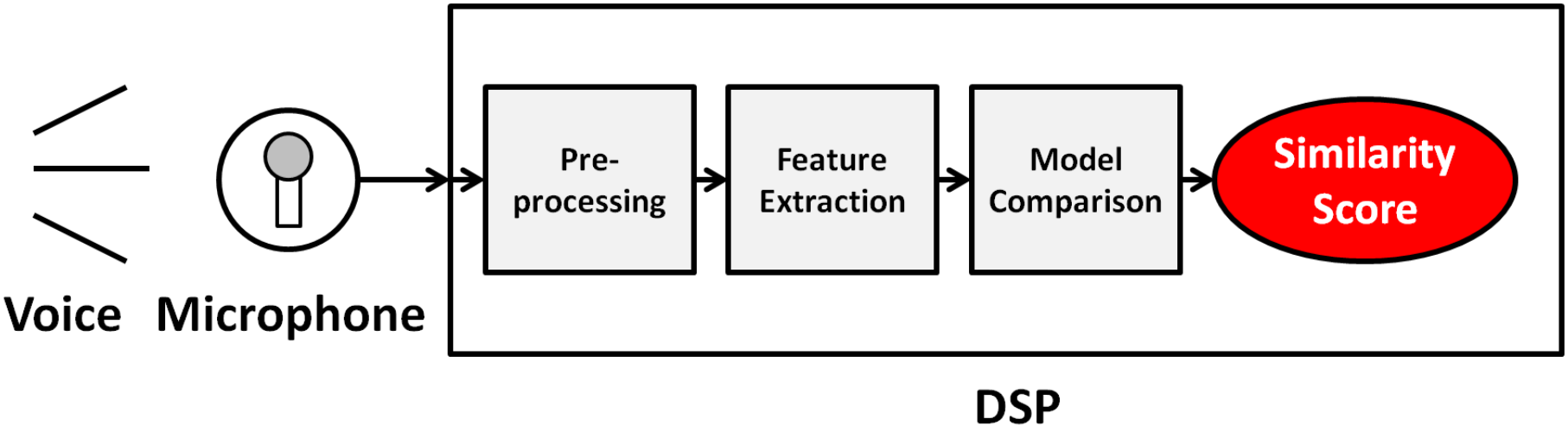


Model Comparison

- ANN training done externally in MATLAB
- Weights updated via back-propagation
- Adaptive learning utilized



Scoring and Decision Making



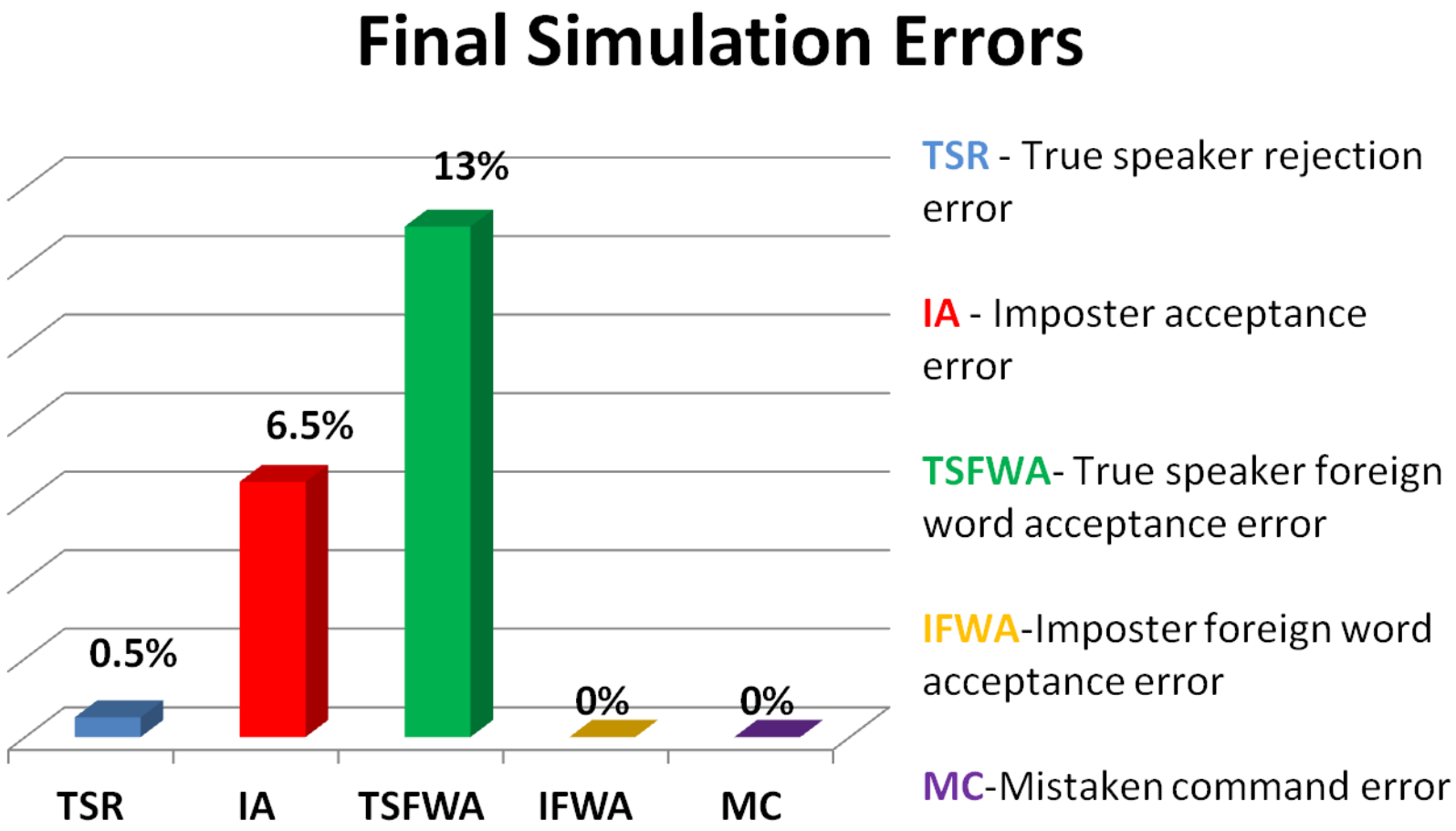
Scoring and Decision Making

- Speech recognition and speaker verification both accomplished with two text-dependent ANN's
- Decision thresholds chosen for optimal error ratios
- Comparing recognition and verification scores yields final command identification

Simulation Conditions

- ANN's trained over 100,000 iterations each
- Training population:
  - 1 true speaker (6 takes) and 11 imposters
  - 2 command words and 4 additional "foreign words"
- Programmed in MATLAB without real time consideration

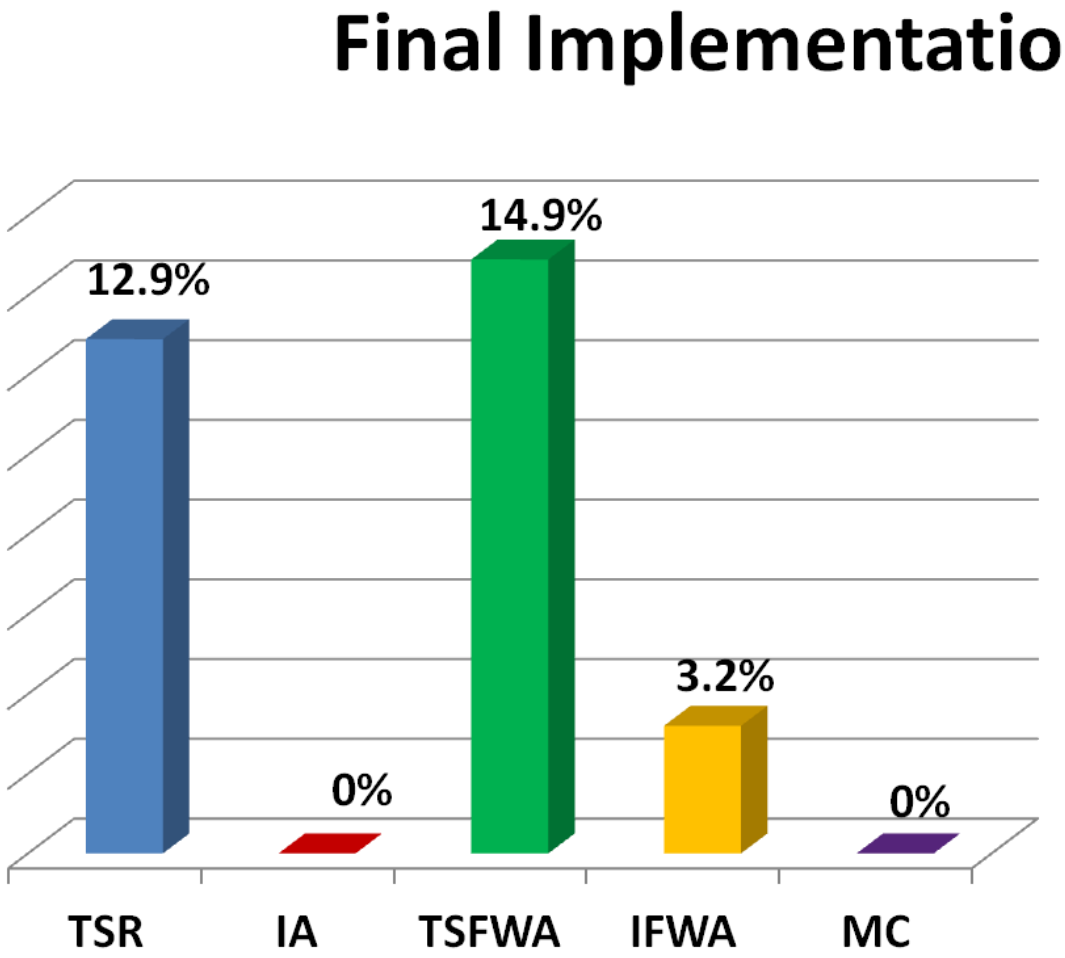
Simulation Results



Implementation Conditions

- ANN's trained over 1,000,000 iterations each
- Training population:
  - 1 true speaker and 5 imposters (10 takes each)
  - 2 command words and 10 "foreign words"
- Programmed in C on a TI eZdsp c5535 development board for real time operation

Implementation Results



TSR - True speaker rejection error

IA - Imposter acceptance error

TSFWA- True speaker foreign word acceptance error

IFWA-Imposter foreign word acceptance error

MC-Mistaken command error

Conclusions

- DSP implementation negatively affects accuracy
- Voice volume and timbre affect accuracy
- Method is poor at rejecting foreign words
- System shows substantial promise

Recap

- Security mechanism
- Methods
  - Windowing
  - MWCC extraction
  - ANN model comparison
  - Decision making
- Simulation results
- Implementation results

References

- [1] J. P. Cambell Jr., “Speaker Recognition: A Tutorial”, NSA, Ft. Mead, MD, Sep. 1997.
- [2] F. K. Soong et al., “A Vector Quantization Approach to Speaker Recognition”, AT&T, Murray Hill, NJ, 1985.
- [3] T. Kinnunen et al., “Comparison of Clustering Algorithms in Speaker Identification”, Univ. of Joensuu, Joensuu, Finland.
- [4] A. K. Jaine et al., “Artificial Neural Networks, A Tutorial”, Michigan State University, East Lansing MI, Mar. 1996.
- [5] Practical Cryptography, “Mel Frequency Cepstral Coefficient (MFCC) Tutorial”, http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/, Oct. 2013.

Autonomous Vehicle Speaker Verification System (AVSVS)

By: Aaron Pfalzgraf and Chris Sullivan
Project Advisor: Dr. Jose Sanchez