Autonomous Vehicle Speaker Verification System

Functional Requirements List and Performance Specifications

Aaron Pfalzgraf

Christopher Sullivan

Project Advisor:

Dr. Jose Sanchez

4 November 2013

Introduction:

This project involves developing an intelligent control system for an autonomous vehicle. The autonomous vehicle is to be controlled by voice commands from the operator, and the system must recognize the difference between an authorized and unauthorized user. A speaker verification system will be used to accomplish this task.

Goals:

- Develop a system that accepts commands only from a specific list of users
- Integrate this system into a speech recognition-based vehicle control system
- Control the vehicle using existing systems from previous senior project
  - Use existing hardware controlled through I2C

System Block Diagram:

Figure 1 shows the overall system block diagram. The operator speaks a command into the microphone. The data from the microphone is passed into the digital signal processing system (DSP). Pre-processing is performed to remove noise; the filtered signal is passed into the feature extraction block to generate a series of feature vectors that describe the signal. These feature vectors are passed into a neural network where a comparison is made between the current audio sample and the known model of the authorized operator. The known model of the operator is determined through a set of training data that is passed into the neural network. The weights of the network are updated using this training data through the back-propagation algorithm. If the current sample is similar enough (within a to be determined threshold), then the system will accept the command from the operator and transmit the corresponding motor control signals to the autonomous vehicle control system. The control system either starts or stops the robot's motor depending on the command given by the user. Figure 2 details the software functionality of the speaker verification system implemented on the DSP.
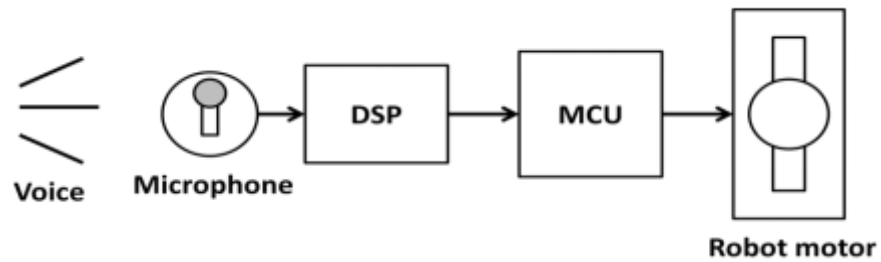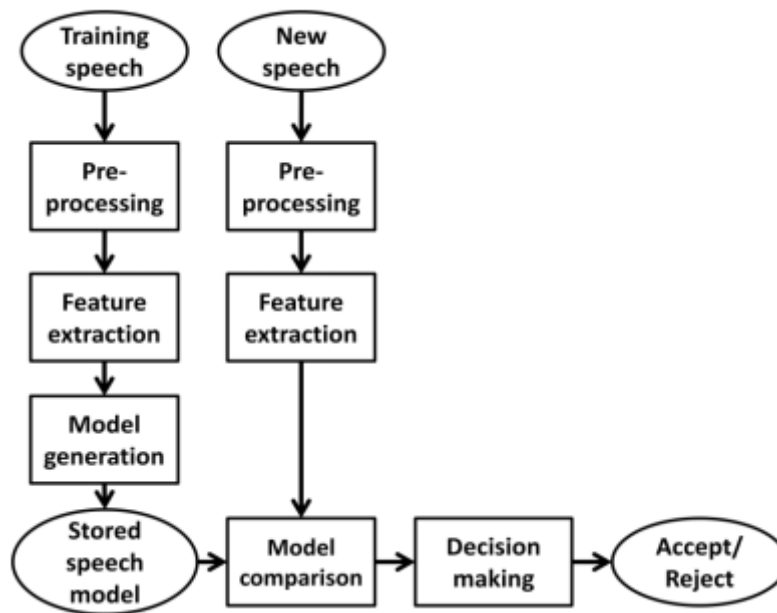
Fig. 1, Hardware Connections

Fig. 2, High level software block diagram

Hardware Requirements:

Microphone

- Omni-directional pick-up pattern for successful voice control regardless of the operator's position in the room
- Signal output via standard 1/8 in. cable, may be accomplished with an adaptor
- Capable of 16 kHz sampling frequency (digital output on microphone)
- Nearly flat 0dB frequency response on the range 100 Hz to 8 kHz
- Passive dynamic microphone, cannot require +48V phantom power (required for active circuitry in some microphones)

Digital Signal Processor

- The DSP used shall be the TI5505 eZdsp

Motor Control System

- The existing autonomous vehicle and MCU from last year's autonomous vehicle senior project shall be used

- Interfacing shall be accomplished with I2C protocol

Software Requirements:

- Basic control commands: STOP, GO, LEFT, RIGHT

- More may be added, such as autonomous routines (e.g. figure 8)

- Speech signal divided into 20 ms - 40 ms frames via Hamming windowing with 33% to 50% overlap

- System shall function properly in a mildly noisy environment

- Maximum operator-to-vehicle distance for proper functionality shall be greater than 10 feet (minimum)

- Operator rejection error shall be minimized to under 1% for safety reasons

- Imposter acceptance error is desired to be under 2% but may be modified to accomplish the desired operator rejection error percentage

- DSP will be programmed in the C language using the Code Composer Studio integrated design environment

- Authorized operator speech models may be generated in MATLAB and hardcoded in C depending on their complexity (online vs. offline training)

- Artificial neural network shall be used to compute the similarity between the current speaker and the authorized operator speech model

- Network shall be trained using the back-propagation algorithm to update the weights of the system

- Feature extraction shall be carried out by Mel cepstral coefficients

- 10 or more coefficients per frame of speech

- Speaker verification shall be accomplished without delay time perceptible to the operator ($t \leq 100ms$)

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.