

The Impact of Microphone Location and Beamforming on In-Vehicle Speech Recognition

Scott Amman, John Huber, Francois Charette, Brigitte richardson, and Joshua Wheeler
Ford Motor Company

ABSTRACT

This paper describes two case studies in which multiple microphone processing (beamforming) and microphone location were evaluated to determine their impact on improving embedded automatic speech recognition (ASR) in a vehicle hands-free environment. While each of these case studies was performed using slightly different evaluation set-ups, some specific and general conclusions can be drawn to help guide engineers in selecting the proper microphone location and configuration in a vehicle for the improvement of ASR. There were some outcomes that were common to both dual microphone solutions. When considering both solutions, neither was equally effective across all background noise sources. Both systems appear to be far more effective for noise conditions in which higher frequency energy is present, such as that due to high levels of wind noise and/or HVAC (heating, ventilation and air conditioning) blower noise. Microphone location was also shown to have a substantial effect on the performance of the ASR system. The results from both studies showed that simply moving a single microphone from the overhead console (OHC) to the sun visor near the driver can provide a great benefit in ASR performance without the cost of multiple beamforming microphones. For the sole purposes of speech enhancement of the driver, it is recommended that moving a single microphone from the OHC to the driver's sun visor position be performed before additional microphones are added. However, beamforming could provide a role if the desire is to include other occupants in a voice session. For instance, a centrally located beamformer could be steered toward the vehicle occupant wishing to issue a voice command. Ultimately, a user-case strategy of the voice control application will determine the microphone configuration and location(s).

CITATION: Amman, S., Huber, J., Charette, F., richardson, B. et al., "The Impact of Microphone Location and Beamforming on In-Vehicle Speech Recognition," *SAE Int. J. Passeng. Cars – Electron. Electr. Syst.* 10(2):2017, doi:10.4271/2017-01-1692.

INTRODUCTION

The use of multiple microphones in the passenger compartment of an automobile can be used to improve the signal (speech) in a noisy environment. Signal processing technologies such as beamforming can take advantage of phase relationships between multiple microphones to help reduce the background noise present in the speech as seen by the Automatic Speech Recognition (ASR) system present in the vehicle. Fundamentally, this technology makes use of the known phase shift or time delay between the microphones when the speech is coming from a known location. Combining the microphone signals, and accounting for the time delay for the known source location, will constructively add the speech signals.

Conversely, sound emanating away from this known location will have a different phase relationship that is not guaranteed to constructively combine. This allows a multiple microphone array to focus on a specified region in the car cabin and reduce the contribution of vehicle noise outside the region. For this application, the region of interest is the talker in the vehicle issuing a speech command to the ASR system. Thus, a beamformer can be used to

improve the signal-to-noise ratio of the speech before it is processed by the ASR system. A good tutorial on microphone arrays can be found in [1].

There are many variations on the previously described methodology. In this paper we examine two case studies in which two different suppliers of microphone array processing have processed multiple microphone speech signals that were captured in a noisy car environment. Even though the two studies were performed using different methodologies, some general conclusions can be made.

SUPPLIER 1 EVALUATION

Data Acquisition

The first investigation using multiple microphones was conducted with the microphone layout as shown in [Figure 1](#). Two microphones configured in a beamforming arrangement were located in the overhead console (OHC) and a single microphone was located just above the sun visor (SV).

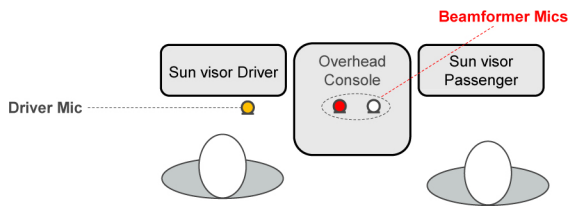


Figure 1. Microphone array used for utterance plus background noise signal collection

The vehicle background road noise was recorded on the test track. Three of the four road noise conditions were 60, 80 and 100 kph on a brushed concrete road surface. The fourth condition was 120 kph on an asphalt surface. HVAC (heating, ventilation and air conditioning) blower was set to off for all of these conditions (Table 1).

Table 1. Noise Conditions for Supplier 1 Evaluation.

Noise Condition		
Surface	Speed (kph)	HVAC Defrost
Brushed Concrete	60	Off
Brushed Concrete	80	Off
Brushed Concrete	100	Off
Asphalt	120	Off

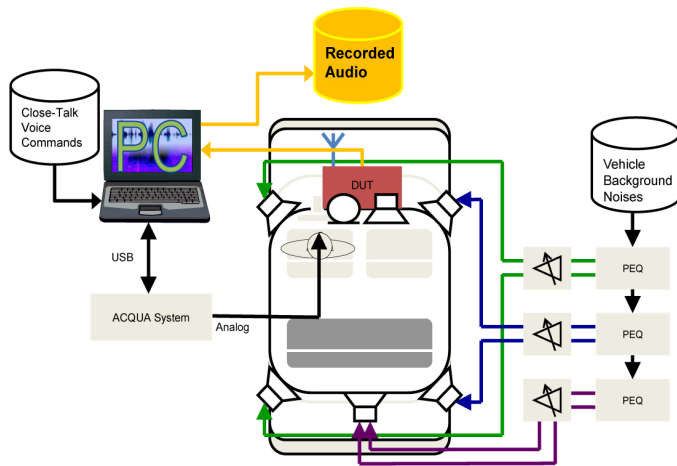


Figure 2. Process for collection of noisy speech utterances.

The voice control grammar (list of allowed commands) utilized, was grammar used in a production vehicle. Speech corpus size was 380 utterances for each of the four driving conditions. The close-talk utterances were played through a head and torso simulator (HATS) located in the driver seat (95th percentile male seating position) while the background road noise was recreated in the vehicle according to the ITU (International Telecommunications Union) protocol defined in [2] for the simulation of recorded vehicle background noise in the car cabin. Figure 2 shows the configuration of the test set-up for noisy speech collection using the Head Acoustics ACQUA speech quality measurement system. The ACQUA system has the ability to play the speech utterances while simultaneously recording the microphone outputs. Once the outputs of the three microphones were

collected, the two microphones for the beamformer arrangement were then post-processed to provide a single channel of noise-reduced speech.

Figure 3 plots the ASR performance for a single microphone in the OHC (one of the beamformer microphones), a beamformer in the OHC, and a single microphone in the sun visor location. In this case, the performance measure used was Command Error Rate (CER %) as reported by the supplier. A command may contain multiple words. If any of the words are not recognized, misrecognition of the entire command will result. The percentage of these missed commands over all issued commands is the CER. It should be noted that absolute values are not necessarily representative of real word results since the Lombard effect was not considered before mixing the speech with the noise. However, relative trends in the results are valid.

This data clearly shows that much more benefit can be gained by moving the microphone to the SV than by adding an additional microphone in the OHC to enable beamforming. Across all conditions, the OHC beamforming configuration provided about a 3.5% absolute improvement in CER over the single microphone located in the OHC. However, moving the single microphone (OHC) to the SV provided an 8.6% improvement in CER. Also, the beamformer had the most impact on the higher speed noise conditions (100 and 120 kph). The reason for this will be discussed later.

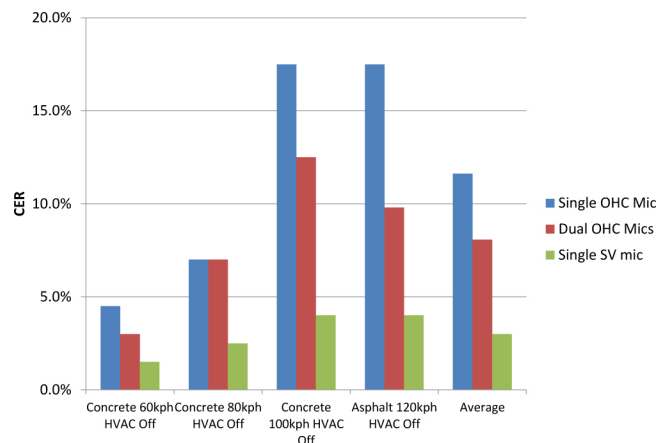


Figure 3. Command Error Rate (%) for microphone configurations.

SUPPLIER 2 EVALUATION

While this study was done under slightly different test conditions and microphone configurations, it will be shown that the conclusions are very similar to the first study.

Data Acquisition

The microphone positioning was reversed from that of the first evaluation. The two microphones used for beamforming were located near the middle of the sun visor. Another single microphone was located in the OHC (see Figure 4). This microphone was used for comparison purposes since many microphones in today's vehicles have microphones located in the OHC.