

Systems Engineering Approach for Voice Recognition in the Car

Joshua Wheeler, Brigitte Richardson, Scott Amman, An Ji, John Huber, and Ranjani Rangarajan
Ford Motor Company

ABSTRACT

In this paper, a systems engineering approach is explored to evaluate the effect of design parameters that contribute to the performance of the embedded Automatic Speech Recognition (ASR) engine in a vehicle. This includes vehicle designs that influence the presence of environmental and HVAC noise, microphone placement strategy, seat position, and cabin material and geometry. Interactions can be analyzed between these factors and dominant influencers identified. Relationships can then be established between ASR engine performance and attribute performance metrics that quantify the link between the two. This helps aid proper target setting and hardware selection to meet the customer satisfaction goals for both teams.

CITATION: Wheeler, J., Richardson, B., Amman, S., Ji, A. et al., "Systems Engineering Approach for Voice Recognition in the Car," *SAE Int. J. Passeng. Cars – Electron. Electr. Syst.* 10(1):2017, doi:10.4271/2017-01-1599.

INTRODUCTION

ASR and Hands Free Communication (HFC) capabilities have become prominent in the automotive industry. With the now common use of mobile personal assistants and smartphones with Bluetooth capability, customer expectations for built-in ASR and HFC systems have increased significantly. As graphically represented in [Figure 1](#), the main systems that impact a customer's satisfaction with their system can be broken into two halves, the "usability & functionality" portion of the system, and the "core automatic speech recognition performance". The success of the latter can be quantified by using the objective metrics of Word Error Rate (WER) and Sentence Error Rate (SER). The voice recognition software/hardware, and elements of the vehicle design, need to be engineered in concert to ensure satisfactory performance of the core ASR system to meet the customer's expectations. It's not enough to design these disparate factors in a vacuum, independent of their influence on the ASR and to one another.

VEHICLE DESIGN SYSTEMS THAT AFFECT VOICE RECOGNITION PERFORMANCE

The success of interpreting a driver's voiced command depends not only on the ASR system hardware and software, but on the design and cooperation of many other parts of the vehicle as well. In the following sections, those major systems are reviewed as single factors, and their effect on the embedded ASR engine's performance quantified, objectively measured by the SER.

As noted, for the purpose of this paper, only single factor analyses are included, but there can certainly be interactions present between these single factors that will cause additional effects (such as seat position and microphone location placement). Those are best explored in a separate, additional discussion.

Background Masking Noise

The ASR system performance is highly dependent on the level of background or "masking" noise that competes with the speech engine's ability to correctly convert the driver's speech to actionable commands. HVAC noise and environmental noise (like road and wind noise) provide high amplitudes of broadband frequency content that affects the signal to noise ratio (SNR) within the vehicle cabin, and masks the user's speech.

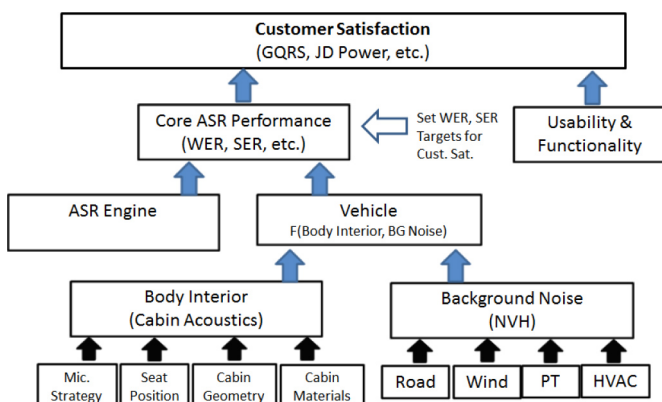


Figure 1. The systems and subsystems which are responsible for the customer's satisfaction with their ASR and HFC system.

Environmental Noise

Road and wind noise levels are highly dependent on design characteristics of the tires, suspension, exterior body shape, and sound package that isolates the cabin from the surrounding environment. Road noise sound pressure levels are mostly dominated by low frequency content that does not contribute in the frequency range where the current generation of automotive speech recognition technology is focused (about 250Hz to 8kHz). However, excessive tire tread sizzle, tire cavitation noise, and transient impact sounds will cause higher frequency issues for the ASR engine. High levels of wind noise caused by aspirations or aerodynamic properties of the exterior body design will cause broadband excitation that also influences the high frequency range where we are trying to preserve the clarity of speech. Figure 2 shows a relationship between escalating environmental noise and the vehicle's SER, as a sedan is driven at higher speeds on a highway-like concrete surface, increasing both the road noise and the wind noise. Unsurprisingly, as the environmental noise increases (shown in decibels, a metric whose characteristics are closely associated with the subjective impression of road noise), the SER also increases in a nearly linear fashion.

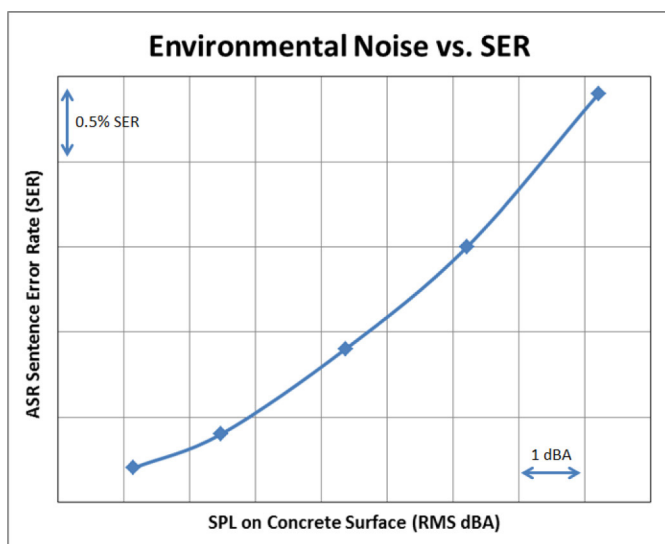


Figure 2. Environmental noise from a sedan driving at various speeds on a concrete surface vs. ASR engine SER. In this example vehicle, for every 1 dBA the environmental noise increases, the SER performance degrades approximately 0.3%.

HVAC Noise

The HVAC system provides a significant source of high frequency energy to mask the driver's speech in the ASR engine. The HVAC duct design, the blower's speed and allowed air flow, and the panel or defroster vent and outlet designs all play a part in the character and amount of HVAC noise present in the cabin. As shown in Figure 3, when the HVAC noise increases, the system's SER increases in a nearly linear fashion. It can also be said from this data that the HVAC loudness (shown here in terms of Sones, a metric whose

characteristics are closely associated with the subjective impression of HVAC noise levels) can be considered to be a more dominant source of ASR engine disturbance than the environmental noise previously discussed.

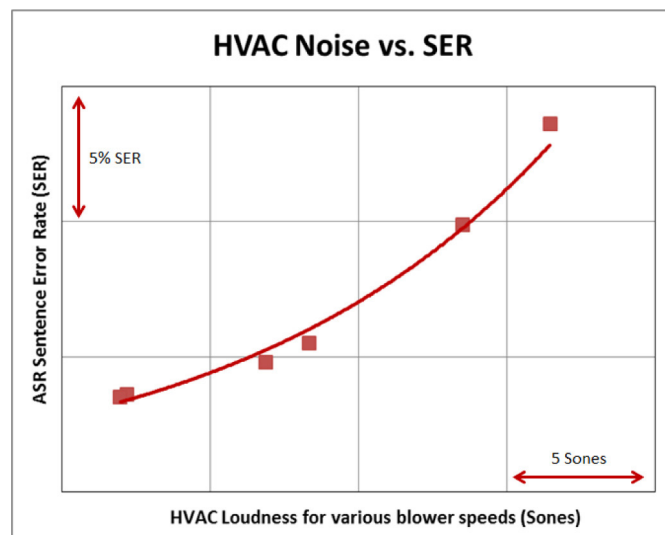


Figure 3. HVAC noise from a sedan at various blower speeds vs. ASR engine SER. For this example vehicle, every 1 Sone the HVAC noise increases, the SER performance degrades approximately 0.6%.

Microphone Location

Proximity to Driver

The placement of the vehicle microphone is of vital importance to ensure the driver's speech is clearly heard above the background noise. The SNR in the cabin is significantly affected by how far away the microphone is packaged from the driver, but competing needs such as studio design, cost, and safety may ultimately place the microphone in a non-optimal location for voice recognition. The majority of automotive manufacturers choose to package the microphone in the overhead console (where it is more easily packaged with other electronic modules), or in the headliner in front of the driver. This second location provides the best SNR for the ASR system, as it places the microphone in the most direct path of the driver's speech. Figure 4 shows an example of how much closer a microphone in the headliner can be to the driver's mouth than when packaged in the overhead console. These actual dimensions were measured in the experiment that generated the data in Figure 5, showing a comparison of the SER in a sedan when a microphone is packaged in either the overhead console, or in the headliner above the sun visor. In addition to the placement of the microphone being important, the orientation of the microphones cardioid pickup pattern can affect the ASR system. To reduce the influence that factor in this test, omnidirectional microphones are used. This data is taken in the presence of defroster noise.