Analysis of Vehicle Voice Recognition Performance in Response to Background Noise and Gender Based Frequency

2017-01-1888
Published 06/05/2017

Rasheed Khan
Hyundai America Technical Center

Mahdi Ali
Hyundai Motor Co.

Eric C. Frank
Bruel & Kjaer Sound/Vib Meas A/S

CITATION: Khan, R., Ali, M., and Frank, E., "Analysis of Vehicle Voice Recognition Performance in Response to Background Noise and Gender Based Frequency," SAE Technical Paper 2017-01-1888, 2017, doi:10.4271/2017-01-1888.

Abstract

Voice Recognition (VR) systems have become an integral part of the infotainment systems in the current automotive industry. However, its recognition rate is impacted by external factors such as vehicle cabin noise, road noise, and internal factors which are a function of the voice engine in the system itself. This paper analyzes the VR performance under the effect of two external factors, vehicle cabin noise and the speakers' speech patterns based on gender. It also compares performance of mid-level sedans from different manufacturers.

Introduction

According to the U.S. Initial Quality Study (IQS), released in June 2016, built-in voice recognition frequently does not recognize or misinterprets commands. This problem has been ranked number as the number one issue in the quality study report [1].

Automatic Speech or Voice Recognition by a machine has been a subject of research for almost four decades. In recent years, enabling hands free communication Voice Recognition (VR) systems have become an integral part of the infotainment systems in the current automotive industry as a form of safety for drivers. A general block diagram of a task oriented voice recognition system is shown in Figure 1 [2] and [3]:

It has become a necessity to have a reliable and consistent voice recognition system in a vehicle that performs to the satisfaction of the customers. Vehicle background noise is thought to be one of the external factors that affect voice recognition performance in a negative way. Many studies and algorithms have performed and proposed to improve VR in vehicular applications. In [4], the author analyzed the applications of voice technology in modern automobiles from two aspects, namely, speech synthesis and speech recognition.

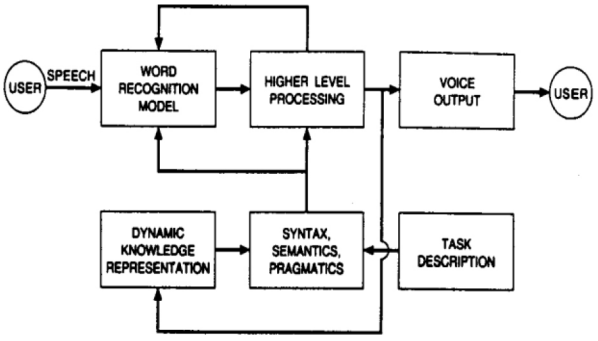

Figure 1. Block diagram of a task oriented voice recognition system

Authors, in many literatures, have proposed methods to improve VR in vehicles. A new constrained switched adaptive beamforming algorithm (CSA-BF) for speech enhancement and recognition in real moving car environments was proposed in [5]. The proposed algorithm consists of a speech/noise constraint section, a speech adaptive beamformer, and a noise adaptive beamformer.

Several literatures have explored different techniques to reduce noise in vehicular VR applications. In [6], several noise reduction algorithms were compared. These algorithms included popular noise reduction techniques such as spectral subtraction and Wiener filtering. Yi Hu and P. Loizou evaluated the performance of several objective measures in terms of predicting the quality of noisy speech enhanced by noise suppression algorithms. These include four classes of speech enhancement algorithms, which are spectral subtractive, subspace, statistical-model based, and Wiener algorithms [7]. Some authors have explored the possibility of using paired microphones for speech recognition systems. This method uses a subtractive microphone array to estimate noise and subtracts them from the noisy speech signal using spectral subtraction [8].

Vehicle voice recognition system performance is impacted by a variety of factors, in addition to background noise, which includes the characteristics of the speaker's speech which may be an effect of their gender. Male and female voice characteristics are inherently different, and some of these variations are investigated in this work. The purpose of this study is to analyze the effect of vehicle background noise and the speaker's voice frequency (based on gender) on the performance of voice recognition. In this paper, three vehicles (herein referred to as Vehicle 1, Vehicle 2, and Vehicle 3) with a similar cabin profile, were tested in five different road conditions (70 mph, 45 mph, Idle HVAC off, Idle HVAC on and vehicle ignition off). The data from the results of this study was analyzed and compared as to how background noise affected the performance of the voice recognition of each vehicle in different driving conditions.

This paper has been organized in various sections to explore the variables which impact the recognition rate in each vehicle. The first section presents an introduction to this analysis. The next section explains the testing setup for all different speakers, phrases, and vehicle conditions. A third section presents a discussion of the results. A lab-based study was designed to identify the impact of speech characteristics in a controlled environment. This and its results are discussed in subsequent sections. Finally, a conclusion section is presented at the end of the paper.

Test Parameters and Strategy

Test Conditions

Three vehicles were tested in 5 different operating conditions:

1. Steady-state 45mph
2. Steady-state 70mph
3. Idle Park HVAC Off (parking lot)
4. Idle Park HVAC On (parking lot)
5. Vehicle Off (parking lot)

Instrumentation

Each vehicle was instrumented with multiple microphones at various locations including at the Voice Recognition (VR) microphone as installed in the vehicle. There were seven microphones in total used to instrument each vehicle:

1. Driver Left Ear (DLE)
2. Driver Right Ear (DRE)
3. Passenger Binaural Head Left (PLE)
4. Passenger Binaural Head Right (PRE)
5. Driver outboard/left corner
6. Driver center
7. Near VR mic location (Driver right for all)

All commands were recorded at all microphones and the time signals were recorded to a computer's hard drive. Background noise during each of the 5 test conditions was also recorded and compared between vehicles.

Another objective of this test was to confirm that the VR microphone was in an optimal location within the cabin. Data from microphones 5, 6, and 7 of the previous list were compared to accomplish this.

Speakers and Script

Voice Recognition evaluations were performed in each of the three vehicles with 6 male and 6 female speakers enunciating the same 20 commands during the various operating conditions. Time domain data for each test was recorded and compared to a log of whether each command was recognized. The twelve speakers were recorded during various operating conditions announcing both mono-syllable (i.e. call mom) and multi-syllable (i.e. call Isabella) commands. Each was performed three times for repeatability. These vehicle-based tests were conducted in parking lots for non-moving conditions and on public roads during conditions for driving events. The drivers were instructed to speak with volume and pronunciation gate as they normally would while using a VR system while driving. They were also directed to face straight forward while issuing their commands to limit the number of variables in the experiment.

A list of the commands used during this test is presented in Table 1.

Table 1. List of Commands used during test

No	Command	No	Command
1	Call Dad	11	Call Liz
2	Call Neal	12	Call Jen
3	Call Joe	13	Call Val
4	Call Mom	14	Call Isabella
5	Call Lynn	15	Call Pablo
6	Call Josh	16	Call Simon
7	Call Kim	17	Call Jesus
8	Call Tom	18	Call Dana
9	Call Ted	19	Call Olivia
10	Call Sean	20	Call Abigail

Results Discussions

Recognition Rate and the Effect of Background Noise

The average overall background noise level for all vehicles is shown below for various operating conditions. Each is the average of three tests. The spectral difference of the vehicles during on-road driving conditions is presented in Figure 2.

Table 2. Average Cabin Noise Level for Different Vehicles (report at Driver's Right Ear)

	Vehicle 1		Vehicle 2		Vehicle 3	
Vehicle Condition	dBA	stdev	dBA	stdev	dBA	stdev
Vehicle Off	36.5	2.2	36.9	0.0	36.4	3.6
Idle, HVAC = Off	44.8	4.2	42.4	0.5	51.3	1.5
Idle, HVAC = On	52.3	0.9	56.8	1.6	53.0	1.3
45mph, constant speed	59.6	2.8	58.9	1.0	61.8	1.5
70mph, constant speed	69.1	1.8	66.6	0.3	67.7	0.9

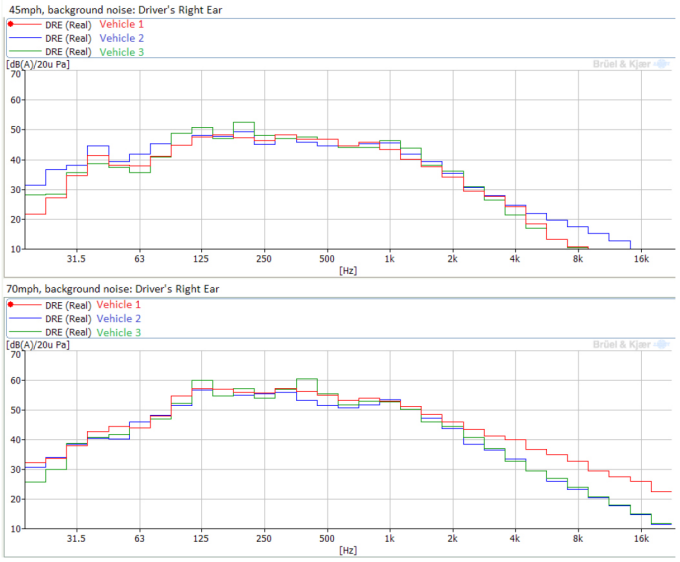

Figure 2. Background Noise Levels at 45mph (top) & 70mph (bottom)

During conditions where the car is not in motion, similar level of background level noise was seen on all vehicles that were tested with the exception at HVAC off in idle condition; Vehicle 3 showed high level in the 125 Hz band.

At 45mph, Vehicle 2 showed the highest noise levels below 100Hz and above 4kHz. At 70mph, Vehicle 1 showed the highest noise levels above 2kHz (most likely attributed to excessive wind noise).

Figures 3, 4, and 5 show the recognition status for each command in each vehicle and operating condition. These values are the total number of successfully recognized commands as a percentage of total trials.

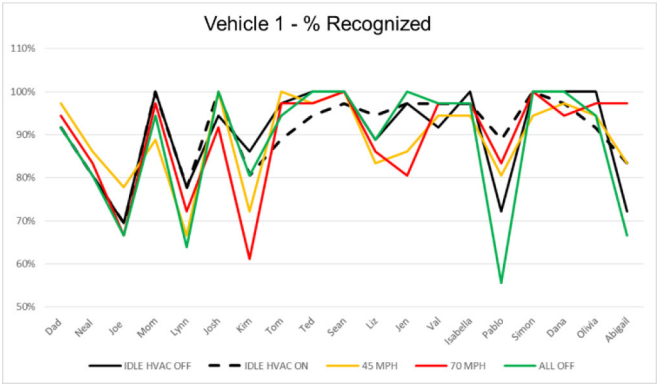

Figure 3. Effect of Background Noise on Vehicle 1

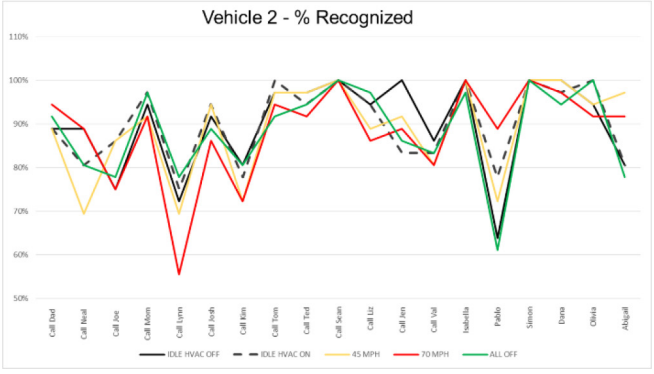

Figure 4. Effect of Background Noise on Vehicle 2

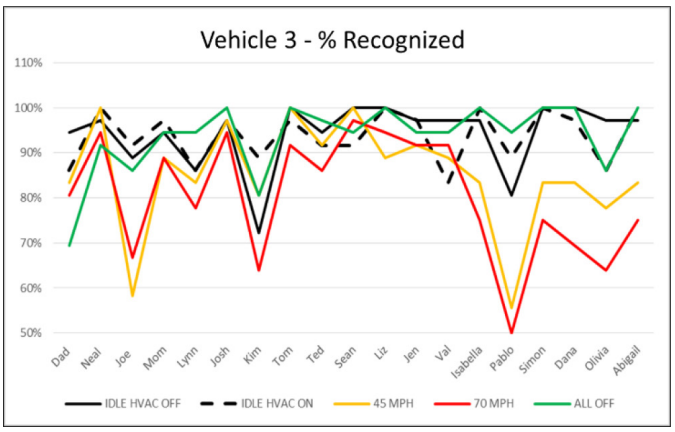

Figure 5. Effect of Background Noise on Vehicle 3

It was observed that one command "Call Pablo" had a poor recognition percentage in all vehicles and all conditions. Additional information about the VR engine and study is required, but it is believed by the authors that this is caused by the frequency and temporal characteristics of this command. The command "Call Jesus" is omitted from these results because it was discovered that some speakers pronounced this command differently.

In Vehicle 3, as shown in Figure 5, it was observed that the recognition rate degraded significantly with a higher background noise level. This is a departure of the recognition rates in the same conditions observed in Vehicle 1 and Vehicle 2 as shown in Figures 3 and 4, respectively. This indicates that the noise cancellation algorithm in Vehicle 1 and Vehicle 2 is more effective in the presence of background noise. At the time of this study, attempts to understand the algorithm in each vehicle were fruitless as OEM's and infotainment providers treat such information as proprietary. This will be the subject of follow-on activities.

Effect of Gender Based Voice Input

Because of the lack of information of the VR engine in each vehicle, mono-syllabic words were the focus of additional analysis of results. This was done under the assumption that poly-syllabic results are more difficult to detect by the VR system and as such, were an additional variable in the experiment.

One of the disparities noted during this analysis was that the recognition rate for the female speakers was significantly lower than that of the male speakers for some commands in Vehicles 1 and 2, but in Vehicle 3, no significant difference was observed. This is displayed in Figures 6, 7, and 8.

In Vehicle 1, the biggest difference between male and female speakers was recognition rate of the name 'Lynn'.

In Vehicle 2, the biggest difference between male and female speakers was recognition rate of the name commands 'Lynn' and 'Neal'

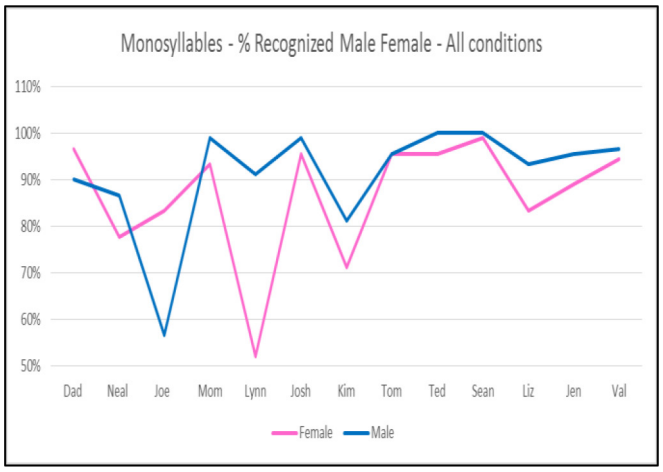

Fig 6. Result of Female Speakers on Vehicle 1

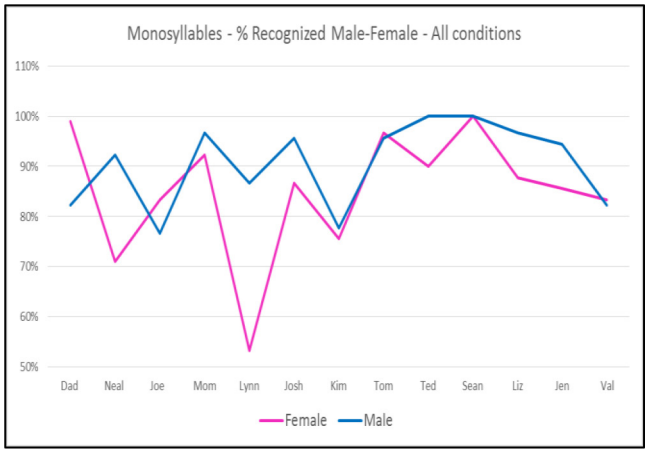

Fig 7. Result of Female Speakers on Vehicle 2

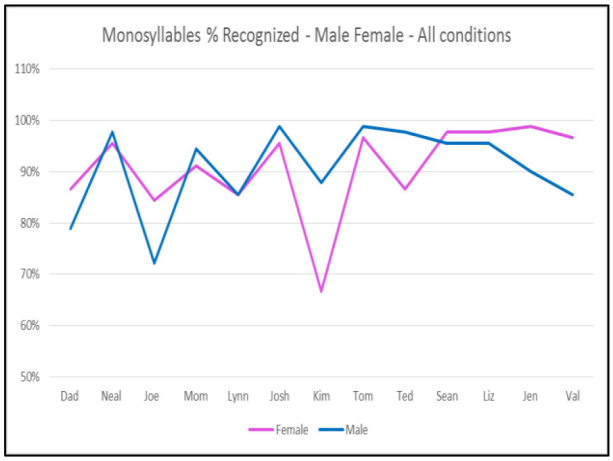

Fig 8. Result of Female Speakers on Vehicle 3

However, the name 'Kim' showed poor recognition rate for the female speakers in Vehicle 3's VR system as Figure 8 shows.

Summary of VR Differences between the Two VR Systems

Overall, the recognition rate in Vehicle 3 is similar to that of Vehicle 1/Vehicle 2, except for the gender based recognition rate. Vehicle 1 and Vehicle 2 exhibit very similar performance which is expected since they are products of the same OEM. The biggest differences between Vehicle 1/Vehicle 2 and Vehicle 3 were the commands "Neal" and "Lynn", with largest difference at 45 mph. Female commands are recognized less than male commands in Vehicle 1 and Vehicle 2, but less difference in Vehicle 3. The commands 'Sean' and 'Simon' are always recognized in Vehicle 1 and Vehicle 2, at all conditions, with all speakers.

Because of the trends of these results, the names "Lynn", "Neal", and "Sean" would be the focus of lab-based studies described below.

Analysis of VR Microphone Position

As stated above, one objective of this study was to determine if the placement of the VR microphone was a variable for consideration in various vehicles, and if so, if the microphone was in an optimal location.

With the vehicle located in a hemi-anechoic chamber, a point source was positioned at the position of a typical driver's mouth. Broadband noise was used to measure a transfer function between the noise source and each response microphone. The results for three vehicles are shown in Figure 9.

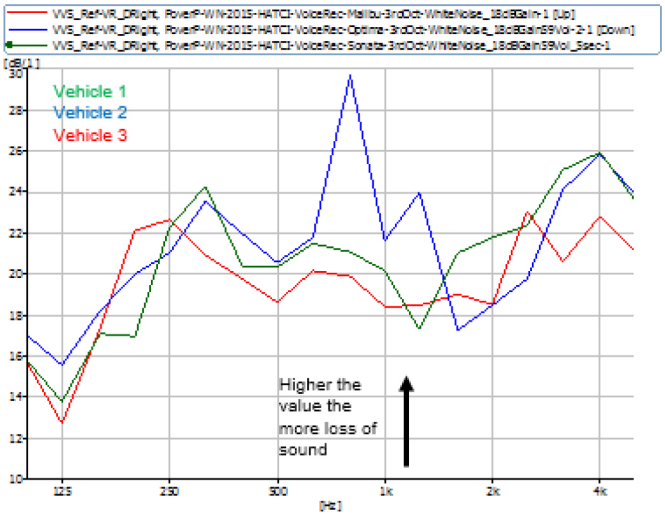

Figure 9. Transmission Loss from driver speech location to VR mic position

This transfer function can be considered the transmission loss at each microphone. The data shows that Vehicle 2 has considerably higher transmission loss at the VR microphone at 800Hz and 1300Hz. These frequency bands are both within the range traditionally associated with typical speech. For Vehicle 2, similar data at the Driver Outboad and Driver Left locations suggest that they would be better suited to measure VR commands from the driver.

Identification of Spectral Differences between Genders

The observation that the recognition rate is consistently different for some commands between male and female speakers warrants investigation into the difference of their speech patterns and frequency content. The figure below shows that analysis of one of the more contested commands between male and female speakers ("Call Neal") shows consistent differences.

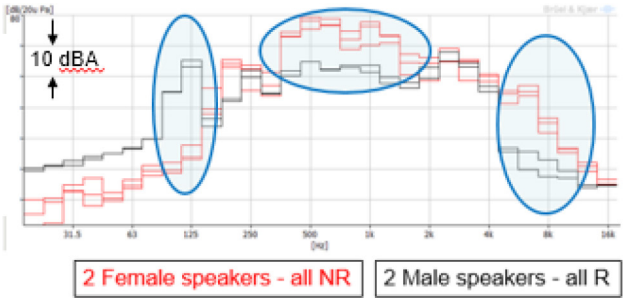

Figure 10. Spectral Difference between Female & Male Speakers for "Call Neal"

When the frequency characteristics for three female speakers (all of which were non-recognized (NR)) are compared to two male speakers (both of which were recognized (R)), there are clear differences in low, medium, and high frequency ranges (100-125Hz, 400-1500Hz, and 5000-8000Hz, respectively).

Generation of Speech Signals & In-Lab In-Vehicle Reproduction

To understand the disparity between the male and female speakers, speech signals were generated to reproduce the recognition rate inside the vehicle in a hemi-anechoic chamber. Commands in these tests were focused on those from drivers which were consistently recognized or consistently not recognized in the on-road tests ("Call Sean", "Call Neal", "Call Lynn"). Hypotheses about the differentiators were formed and the need was identified to test these in a controlled environment.

To control this experiment, free field recordings from each participant were played through a noise source which was position inside the vehicle at the driver's mouth position. The VR system was activated, and then recordings were presented from the noise source to observe the effect on recognition. The baseline recordings were used, as well as others that were synthesized to study the effect of parameters such as relative amplitude of the spoken words during the command, the duration of the words in the command, and the frequency characteristics of the speaker's voice.

The following tasks were undertaken to identify the sensitivity to these variables:

Initial Setup: Validation of Test Procedure

Playing back "Call Sean", recorded in hemi-anechoic chamber by the most recognized male speaker (Bret), to confirm reproducibility of recognition percentage. When a recording was played back through this lab-based system, the results were identical. This gave the team confidence that the artificially-created system could be used to test the sensitivity to variables as described below.

Step 1. Effect of level of "Call Command"
Starting with "Call" and "Sean" at identical level and changing their level between 30 dBA and 80 dBA at 5 dBA increments. This would test the dynamic operating range of VR system

Step 2. Effect of relative level (ΔL) between "Call" and "Command" (Speech Modulation Depth)

Step 3. Effect of interval duration (Δt) between "Call" and "Command"

Step 4. Effect of duration of "Call" & "Command"

Step 5. Effect of frequency composition of female vs. male voice (Example: Neal, as described above).

Summary of Test Signals & In-Lab In-Vehicle Reproduction

Table 3. Results of Lab-based Variable Analysis

Variable Tested	Description of Test Signals	Result
Amplitude of Command	Baseline was a known recognized command The same signal was played back at various levels from 30dBA to 80dBA (in 5dBA increments)	Command is not recognized when played below 55dBA
Relative Amplitude between "Call" & <Command>	"Call Lynne", "Call Sean", "Call Neal" were modified to have different amplitudes (low, medium, high) for each word	Male-spoken commands were very consistently recognized. Recognition of Female-spoken commands was much more variable
Time delay between "Call" & <Command>	The time delay between the two words was modified based on real-life differences. Short, medium, and long gaps were tested	Male-spoken commands were greatly recognized, regardless of time gap. Female-spoken commands were all unrecognized for longest gap
Duration of "Call" & <Command>	Audio-modification software was used to change the duration of each word. Short, medium, and long durations of each word were tested	Male-spoken commands ranged in recognition from always to sometimes recognized. Female-spoken commands were only sometimes recognized
Frequency Filters of "Call <Command>"	Commands were modified based on spectral differences in low, medium, and high frequency ranges as described above.	Sensitivity to relative differences between mid and high frequency ranges (see below)

When modifying the frequency distribution of various commands, the results show that the VR engine is sensitive to these changes and vary depending on the command tested.

In the case of "Sean", when the mid frequency range is amplified causing a larger difference to the high frequency range, it seems more likely not to detect "Sean". In the case of "Neal", when the high frequency range is attenuated creating a larger delta with the mid frequency range, it seems more likely to detect "Neal". Adversely, when the mid-frequency was reduced for "Neal" forming a smaller delta with the high frequency, it seems less likely to detect.

Different words have different frequency content, so the same pattern is not expected to be seen for all commands. It seems there is sensitivity between the relative level of the mid and high frequencies.

Conclusions

Recognition rate for female speakers is poor in both the Vehicle 1 and Vehicle 2 when compared to that of male speakers for some commands. This pattern was not observed in Vehicle 3. Although all vehicles implement the narrow band vocoder technology, the Vehicle 3 VR performance showed less difference in recognition rates between genders.

VR performance generally degrades in the Vehicle 3 with increased background noise, but in Vehicle 1 and Vehicle 2 the performance is not affected by the noise increase. Vehicle 3 rates better in perceived quality in the sense that all participants preferred it due to the ease of usability (not many input requirements to complete a function), despite the degraded VR performance of Vehicle 3. This observation needs to be considered when interpreting results from JD Power and other such surveys.

While the background noise levels are not the highest for Vehicle 3 during driving conditions, its Voice Recognition rate is lower than the other vehicles. This suggests that other variables impact the performance of the VR system.

The laboratory tests show that under ideal and controlled conditions, attributes of the user's speech pattern contribute to the successful use of a vehicle's Voice Recognition system.

While these attributes have been identified in the body of this papers as contributors, full understanding of the VR engine is necessary to optimize the in-vehicle performance and hence, driver experience. In the case of the vehicles studied here, collaboration with the vehicle OEM and infotainment supplier are to be considered as future work.

References

1. jdpower.com, J.D. Power U.S. Initial Quality Study (IQS), EXECUTIVE SUMMARY, June 2016.
2. Rabiner L., Juang B.H., "Fundamentals of Speech Recognition", 1993.
3. Bellestri C., *Automatic applications of voice recognition technology*, 1986
4. Awad, S. "Voice Technology in the Instrumentation of the Automobile", IEEE--Trans. on Instrumentation and Measurement; vol. 37, No. 4; Dec. 1998; pp. 586-590.
5. Zhang X. and Hansen J.H.L., "CSA-BF: A constrained switched adaptive beamformer for speech enhancement and recognition in real car environments", *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 733-745.
6. Prodeus A., "Performance Measures of Noise Reduction Algorithms in Voice Control Channels of UAVs", Proc. of IEEE 3rd International Conference "Actual Problems of Unmanned Aerial Vehicles Developments (APUAVD)", pp. 189-192.
7. Hu Y. and Loizou P., "Evaluation of objective quality measures for speech enhancement", *IEEE Transactions on Speech and Audio Processing*, vol. 16, pp. 229-238, 2008
8. Mizumachi M. and Akagi M., "Noise reduction by paired-microphones using spectral subtraction," Proc. Intl. Conf. on Acoust., Speech and Signal Process. (ICASSP), Vol. II, pp. 1001-1004, 1998.

Contact Information

Rasheed Khan
Sr Engineer, Product Quality
HATCI Electronic Systems Development
6800 Geddes Rd, Superior Twp., MI 48198
Office : (734) 337-2221
rkhan@hatci.com

Mahdi Ali
HATCI Electronic Systems Development
6800 Geddes Rd, Superior Twp., MI 48198
Office : (734) 337-2386
mali@hatci.com

Eric C. Frank
Operations Manager / Senior Project Engineer
Brüel & Kjær - Global Engineering Services
6855 Commerce Blvd, Canton, MI 48187
Mobile: (248) 207-1357
eric.frank@bksv.com

Acknowledgments

The authors would like to thank the Management teams at Hyundai and Brüel & Kjær for their support in the conception and successful completion of this study. Great appreciation is also due to Bahare Naimipour and Valerie Schnabelrauch of Sound Answers, Inc. for their efforts in performing vehicle measurements with great attention to detail.

ISSN 0148-7191

http://papers.sae.org/2017-01-1888