www.kardome.com
info@kardome.com

IN-CAR SPEECH RECOGNITION

Introduction

A natural voice user interface (VUI) allows the user to interact with a system through voice or speech commands and is expected to recognize natural language in acoustically challenging environments. Current speech recognition systems offer an accurate detection and recognition performance for a single speaker in a quiet environment. However, interfering signals, such as a car radio, simultaneous speakers, and environmental noise, negatively affect their performance. Therefore, modern VUIs use sophisticated microphones to focus on the desired speaker, while reducing the effect of interfering speech and the environmental noise.

Kardome™ is an innovative microphone that provides a **spatial focus** on a desired speaker in acoustically challenging environments such as vehicles. Unlike other microphones, which are based mostly on directional beamformers, Kardome™ provides a location-based speech enhancer, which focuses the microphone **on the location of the desired source** rather than toward the direction of the source.

This white paper presents an experimental study that compared the speech recognition rate (SRR) obtained by the Google Speech to Text (GSST) engine in a car travelling at 100 kph with three types of microphones: (i) a single microphone, (ii) directional microphone, (iii) Kardome's location-based microphone.

Experimental Setup

A Skoda Octavia car was equipped with three microphones: (i) a single microphone, (ii) a directional microphone, and (iii) Kardome's location-based microphone. All were installed in the overhead compartment. Three speakers were installed, one each on the headrest of the front passenger, rear left passenger, and rear right passenger. The desired speech signal was composed of 20 short sentences, ~ 6 seconds long each, and the interfering speech signal was a continuous speech recording. Each of the signals was spoken by a native English speaker and recorded using high-quality audio recording equipment in a quiet room.

The considered three scenarios are described in Table 1. In each scenario, the signals were recorded simultaneously by the three microphones, while the car was travelling at 100 kph. The resulting SRR was evaluated by applying the GSST to the recorded signals.

In the first and second scenarios, the desired speech signal was corrupted by a spatial noise induced by the car. In the third scenario, in addition to the spatial noise, an interfering speech signal was delivered to the medium by the rear-left speaker, simultaneously with the desired speech signal, which was produced by the front speaker.

Kardome's location-based microphone was calibrated to pick up the desired speech in all the scenarios.

The directional microphone was calibrated toward the direction of the desired speaker and an appropriate null was steered toward the direction of the interfering speech in the third scenario.

Both methods were set to minimize the output noise level, while the power spectral density of the noise was estimated during periods when speech was absent.

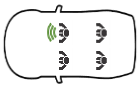
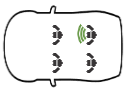
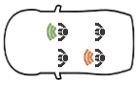
Scenario	Front Speaker	Rear-Right Speaker	Rear-Left Speaker	
i	Desired speech	Off	Off	
ii	Off	Desired speech	Off	
iii	Desired speech	Off	Interfering speech	

Table 1 – Considered scenarios

The input signal to interference ratio (SIR), i.e., the ratio between the desired and interfering signals' power, as measured by the microphones, was set to ~ 10 dB.

Results

The SRR performance for each of the considered scenarios is shown in Figures 1-3. Each bin in the top plot represents the SRR obtained by each of the three microphones for each of the 20 sentences. The blue bins correspond to the single microphone, the red bins to the directional microphone, and the green bins to the Kardome location-based microphone.

The confidence that each sentence was labeled correctly is depicted in the bottom plot in each figure.

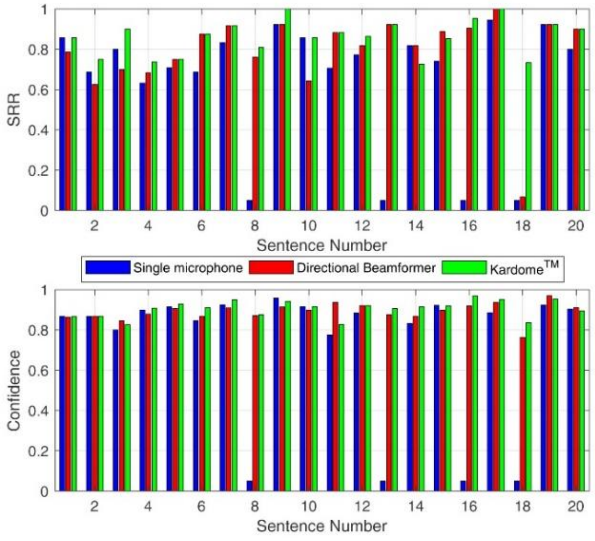


Figure 1 - Google Speech to Text results for the first scenario

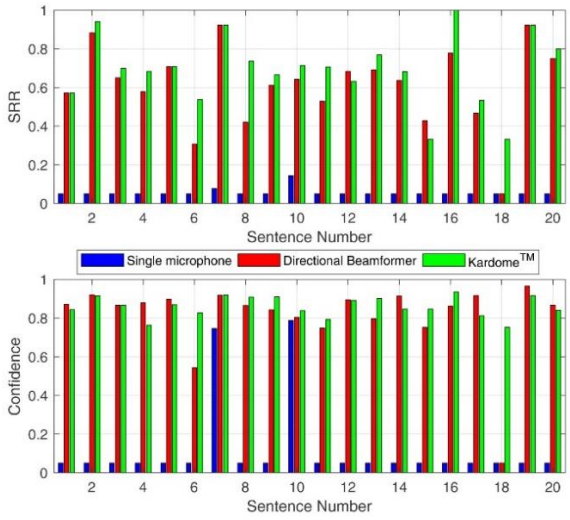


Figure 2 - Google Speech to Text *results for the second scenario*

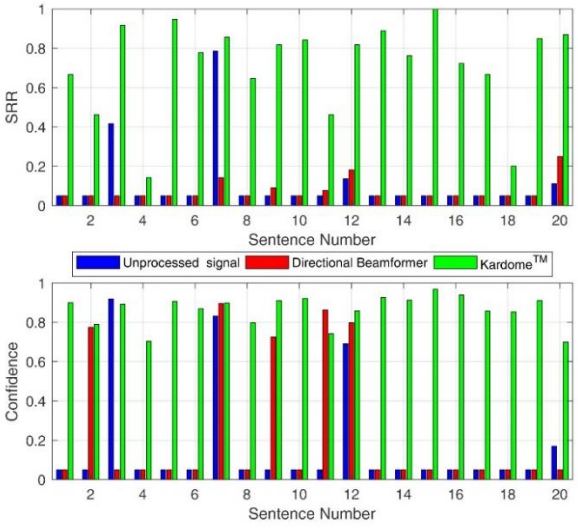


Figure 3 - Google Speech to Text results for the third scenario

Conclusions

The application of speech enhancement methods in a car improved the SRR in all the scenarios. In the absence of interfering speech, the Kardome™ results are slightly better than those of the directional microphone. In the presence of an interfering speech signal, the results of the directional microphone are significantly degraded, whereas a meaningful SRR performance is obtained by applying Kardome™.

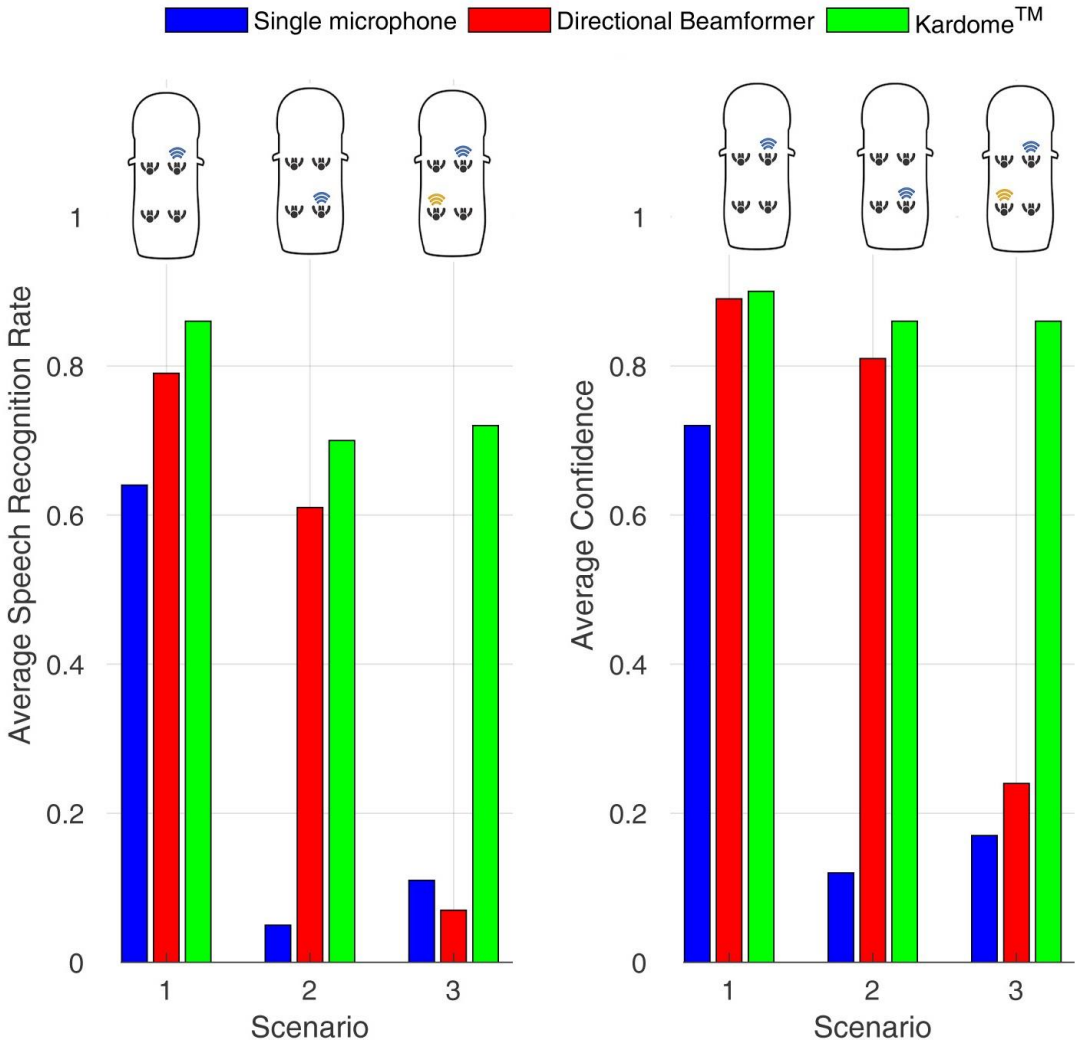


Figure 4 - Average Google Speech to Text results