Microphone Array Processing for Speech Enhancement in Car Environment by Improving Generalized Sidelobe Canceller

Sachin Bainur
MSc. in Automotive Software Engineering
Chemnitz University of Technology
Chemnitz, Germany
sachin.bainur@s2019.tu-chemnitz.de

***Abstract*—Microphone arrays are widely used in speech enhancement algorithms to build smart communication interfaces. These interfaces are used in automobiles, telecommunication, and hearing devices. In this report, improvisation to the algorithm of Generalized Sidelobe Canceller(GSC) is proposed along with usage of Micro-Electro-Mechanical System(MEMS) microphone array to remove the commotion obstruction in-vehicle environment. In the first place, low-frequency concealment is introduced to GSC, it used source area to improve the Blocking Matrix(BM), in this manner adequately decreasing the speech loss due to noise. At last, the denoising model is further improved by using a spectral subtraction algorithm to restore the magnitude spectrum of the signal in additive noise.**

***Index Terms*—MEMS microphones, Denoising model, Generalized sidelobe canceller, Spectral substraction, Microphone array, Frequency, Spectrogram**

I. Introduction

Today speech is developed into an interface for controlling electronic gadgets. Speech recognition has been applied to different applications on account of its effortlessness and lucidity. In recent years speech-based interactive systems are gaining wider applications in automobiles. According to Capgemini reports 73% of drivers will use in-car speech system by the year 2023 [4].

Microphone is a primary source of input to any speech enabled devices, but trivial microphone works well only in relatively high Signal-to-Noise Ratio (SNR) environment i.e. SNR $\geq$ 15dB. Ambient noise in vehicle environment is much higher and consists of strong high and low frequency components resulting in failure of trivial microphones.

Designing MEMS microphones have started a new era in voice applications. Since the year 2014 MEMS microphone has taken over Electret Condenser Microphone(ECM) as being the most popular choice in speech applications in harsh environmental conditions. The advantages of MEMS over ECM are

- ECMs can not tolerate the high temperatures of a reflow soldering operation, so they should always be fastened to a board by hand. The MEMS microphones can be collected as different ICs on a PCB during the equivalent reflow soldering process [5].
- MEMS microphone performance density is much higher. It may be outperformed by a MEMS microphone smaller in size to ECM in a given condition.
- Multiple power modes in MEMS is suitable for speech application, thus reducing power consumption by 90 percentage.
- MEMS microphones provide an easy interface to other devices compared to ECMs.
- Piezoelectric MEMS microphones are dust and water resistant
- MEMS microphone's performance doesn't degrade over years unlike ECMs

TABLE I
Comparison of MEMS and ECM microphones[5][6]

Parameter	MEMS Microphones	Electret Condenser Microphones
Size	2.5*3.35*0.88 mm	9*4*7.05 mm
Sensitivity	5.0 mv/Pa	17.8 mV/Pa
Operating temperature	-40 to +85 °C	-20 to +70 °C
Maximum pressure	172 dB	141 dB
Sensitivity drift with temperature	+/-0.5 dB	+/-4 dB
Bandwidth	10kHz-230kHz	60kHz-200kHz
Diaphragm thickness	<1 μm	>1.5 μm

Table I shows the value-based comparison of different parameters of both MEMS and ECM microphones

Speech assistant in automobiles uses microphone arrays for processing voice. Microphone array beamforming is an effective method of speech enhancement and GSC is a generalized model of the adaptive beamformer. The recent improvements to GSC are based on the presumption that the acoustic source is situated in the most distant field of the array but the reality is acoustic source is in near field array in a real driving environment and incident waves are spherical. By considering this an effective blocking matrix can be generated using the position of the source. Furthermore, the usage of multi-band spectral subtraction to this model minimizes noise efficiently.

The rest of the report is sorted as follows. Design overview of MEMS microphone is given in Section II. Acoustic propagation in driving conditions and concepts of microphone array is explained in Section III. Improvements to GSC are provided in Section IV. Improvements performance and experimental results are shown in Section V. Finally, Section VI gives a conclusion.

II. Design overview of MEMS Microphones

Over the past few years, silicon technology has advanced. A typical design incorporates a MEMS sensor with an Application-Specific Integrated Circuit (ASIC) as shown in Fig. 1. The MEMS microphone is designed using MEMS technology. The sensor has an acoustic pressure-driven membrane that produces an electrical signal for analog microphones amplified or processed by the ADC on ASIC for digital microphones.

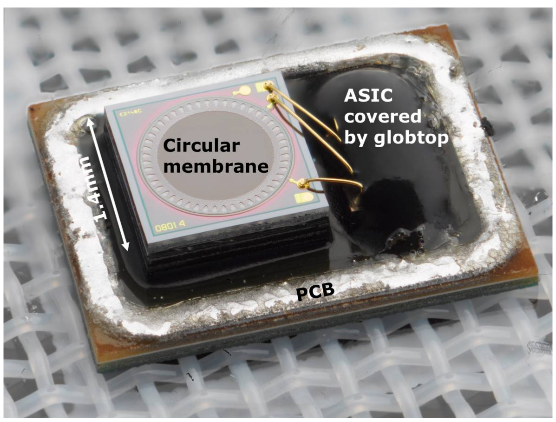

Fig. 1. The typical MEMS membrane with ASIC[2]

MEMS microphone chip is important in transforming acoustic pressure into a capacitive signal. It is primarily composed of a conforming diaphragm and a rigid perforated backplate, which are separated dielectrically from a certain air gap given their different embodiments in material, configuration and manufacture. The diaphragm of the MEMS microphone chip is primarily considered with regard to its direct impact on performance and reliability. Fig. 2 displays typical diaphragm designs. For most commercially available MEMS microphones on the market, type (a) and (b) diaphragms can be seen widely. Type (a) is the best design for reliability to the impact proof because it is free of side constraints to easily achieve efficiency and to prevent failure due to excessive stress. Type (c) instead is weak for impact evidence in its supporting springs. Type (d) is a workable architectural definition.

As shown in Fig. 3, MEMS microphone is also possible to combine diaphragm with perforated backplate. The diaphragm in this configuration move with the attached perforated plates, producing a capacitive change to the conductive substrate [2].

Parasitic capacitance should always be minimized when designing microphones. The structural portions that are not

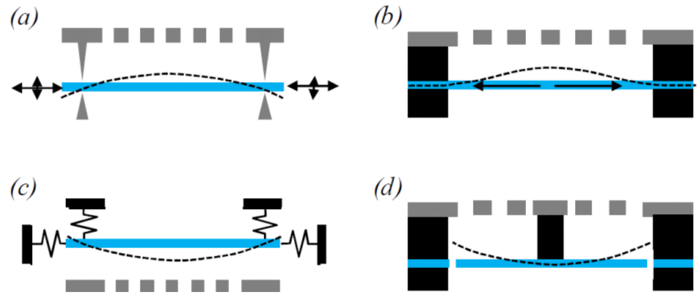

Fig. 2. Typical diaphragm designs in MEMS microphone. (a) Free floating plate (b) Tensile stressed plate (c) Spring supported plate (d) Center-supported plate[2]

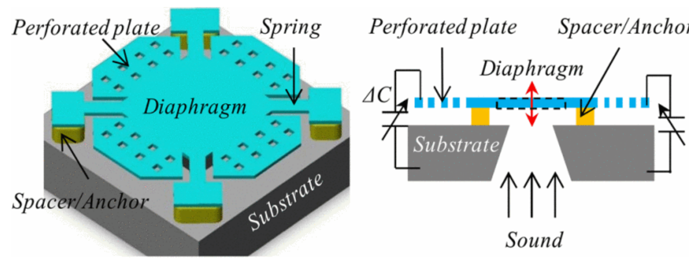

Fig. 3. MEMS microphone without dedicated backplate[2]

moving, such as the wire fasteners and anchors form parasitic capacitance not contributing that reduces the sensitivity of the microphone and the SNR. As illustrated in Fig. 4, an effective way of reducing it is to introduce protective layers with a tendency to shield the areas of the parasite capacitors with the same voltage potential.

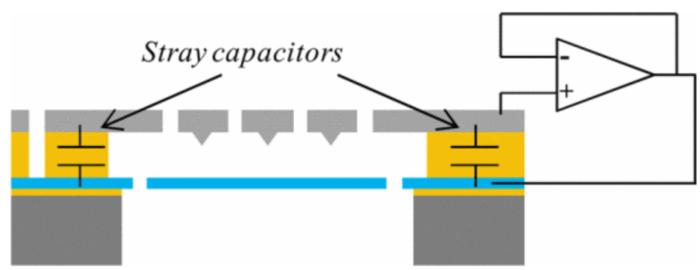

Fig. 4. Guarding layer to shield away parasitic[2]

MEMS microphones with omnidirectional direction capture sounds from any direction equally. Unfortunately, besides the sounds of interest, the sound converted into electric signals also produce unwanted noises. It allows MEMS microphone arrays and the accompanying electronics to be used to enhance the quality of sound. These arrays generate a beam-forging directional reaction that channels unwanted noise in a desired direction and filter out required signals.

III. Acoustic Propagation and Concepts of Microphone Array

For building a MEMS microphone array, two or more microphones are used for audio sounds and for producing the electric signal of each microphone. Electronic circuits are used for the treatment of each microphone's signals before the signals are combined (amplification, delays, filters, etc). Electrical processing improves the desired signals and reduces the unwanted signals. The microphones in the array must either have strictly matched performance requirements, or have to be characterized individually for specification performance in order to be effective in signal processing. The microphone sensitivity is the key parameter that must be balanced correctly in the series. MEMS microphones are conveniently available with closely matched sensitivity tolerances due to semiconductors manufacturing process, making them a good choice for microphone arrays. A typical array arrangement in cars is as in Fig. 5 with driver as source.

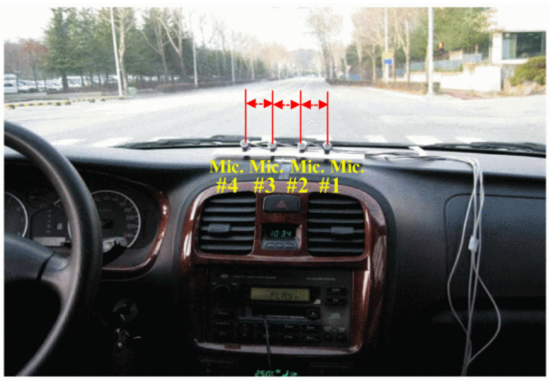

Fig. 5. Microphone array arrangement in car environment[14]

Most of the microphone arrays in speech recognition systems support the idea that acoustic source is in far-field of the microphone array. But in a typical driving environment, the acoustic source is within the near field of array. Hence, incident sound waves are considered as spherical waves [6][7].

For our model, we are using a uniform linear microphone array. Let r be the distance between source and microphone we refer to, θ and φ represent the incidence and azimuth angles respectively. The equation can be derived for spherical wavelet as

$$\frac{1}{r^2}\frac{\partial}{\partial r}(r^2\frac{\partial s}{\partial r})+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\varphi}(\sin\varphi\frac{\partial s}{\partial\varphi})+\frac{1}{r^2\sin^2\varphi}=\frac{1}{c^2}\frac{\partial^2 s}{\partial t^2} \quad (1)$$

In general, acoustic field of a source in near field of array is isotropic hence there's no change in direction of θ and φ. We also know that c=343m/s. Equation (1) can be simplified now as

$$\frac{1}{r^2}\frac{\partial}{\partial r}(r^2\frac{\partial s}{\partial r})=\frac{1}{c^2}\frac{\partial^2 s}{\partial t^2} \quad (2)$$

The solution of wave equation can be written as

$$s(r,w,t)=\frac{A}{r}\exp\{j(wt-kr)\} \quad k=\frac{w}{c} \quad (3)$$

Assuming that there are N elements to receive signals and M signal sources within the area. The reaction of a m^{th} source on the n^{th} array element can be written as

$$x_{mn}(t)=\frac{1}{r_{mn}}s_m(t-\frac{r_{mn}}{c}) \quad (4)$$

Applying Fourier transform on both sides

$$X_{mn}(t)=\frac{1}{r_{mn}}\exp\{-j\omega\frac{r_{mn}}{c}\}\cdot S_m(\omega) \quad (5)$$

Let N(t) be the background noise received by the n^{th} microphone. Therefore, the array response of the n^{th} microphone is

$$y_n(t)=\sum_{m=1}^{M}x_{mn}(t)+n_n(t) \quad (6)$$

$$Y_n(\omega)=\sum_{m=1}^{M}X_{mn}(\omega)+N_n(\omega)$$
$$=\sum_{m=1}^{M}\frac{1}{r_{mn}}\exp\{-j\omega\frac{r_{mn}}{c}\}S_m(\omega)+N_n(\omega) \quad (7)$$

The response transfer function matrix G(ω) of microphone array can be expressed as

$$G(\omega)=\begin{bmatrix}\frac{1}{r_{11}}\exp\{-j\omega\frac{r_{11}}{c}\} & \cdots & \frac{1}{r_{M1}}\exp\{-j\omega\frac{r_{M1}}{c}\}\\ \vdots & \ddots & \vdots\\ \frac{1}{r_{1N}}\exp\{-j\omega\frac{r_{1N}}{c}\} & \cdots & \frac{1}{r_{MN}}\exp\{-j\omega\frac{r_{MN}}{c}\}\end{bmatrix} \quad (8)$$

Considering the Fourier transforms of the signal source, noise and array received signal as S(ω) , N(ω) , Y(ω) respectively.

$$S(\omega)=[S_1(\omega),\cdots,S_M(\omega)]^T \quad (9)$$
$$N(\omega)=[N_1(\omega),\cdots,N_M(\omega)]^T \quad (10)$$
$$Y(\omega)=[Y_1(\omega),\cdots,Y_M(\omega)]^T \quad (11)$$
$$Y(\omega)=G(\omega)\cdot S(\omega)+N(\omega) \quad (12)$$

Since our microphone array is uniform linear, the source of the signal, is projected in two dimensions plane as shown in Fig. 6 where the array is organized

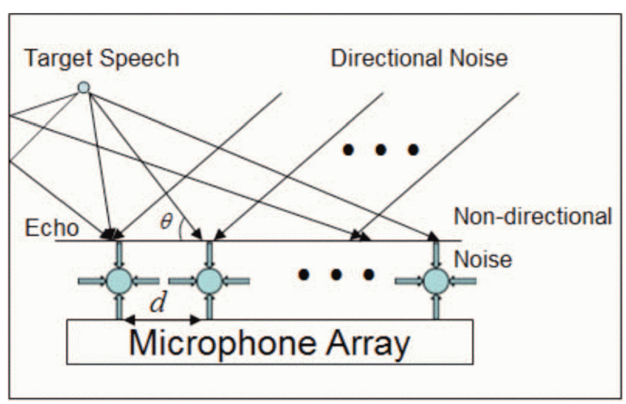

Fig. 6. Uniform linear microphone array[1]

The first element of the array is the reference element, and then the distance from the source of the signal to reference dimension is r_{m1}

$$r_{mn} = \sqrt{(r_{m1}\sin\theta)^2 + (r_{m1}\cos\theta + (n-1)d)^2} \qquad (13)$$

In formula (8), the equation (13) is substituted for obtaining the linear uniform microphone response function matrix for the condition of near-field configuration. Therefore, linear array framework of received signal is

$$Y(\omega) = G_l(\omega) \cdot S(\omega) + N(\omega) \qquad (14)$$

IV. Improvements to GSC

Microphone arrays use beamforming techniques to prioritize signals from particular directions and eliminate noise from other directions. The GSC is an effective model used widely in microphone arrays for beamforming. It consists of multiple channels as shown in Fig. 7. The primary channel is fixed beamformer, shapes a beam in the direction of the look such that the desired voice signal is transmitted and all other signals are suppressed. It mainly used the Delay and Sum Beamforming (DBS) algorithm to perform this task. The secondary channel i.e. auxiliary channel consists of 2 subparts namely a Blocking Matrix (BM) to measure an approximation for a noise source signal by blocking the components of the desired speech signal and an adaptive interface canceller whose value coefficients are used to calculate noise in delay-sum beamforming result. Thus, an enhanced signal can be calculated by eliminating reference noise from delay and sum beamforming results.

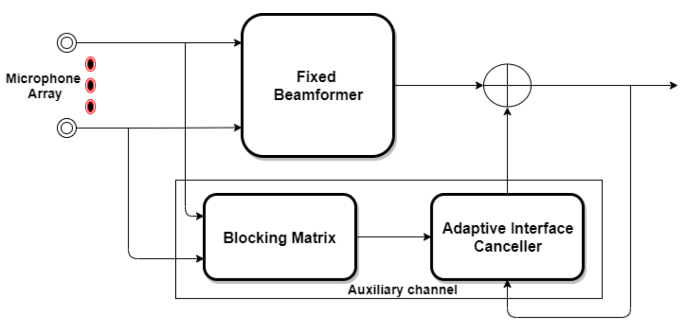

Fig. 7. Block diagram of generalized sidelobe canceller[1]

Though GSC has wide applications and used in most of microphone arrays it has some disadvantages which can be improved by doing few changes.

- Addition of a low frequency suppression
- Blocking matrix improvement
- Using multi-band spectral subtraction algorithm(SSA)

A. GSC with Low-Frequency Suppression

The noise in the driving environment can primarily be divided into coherent and non-coherent noise. The coherent noise can be a conversation between passengers and non-coherent noise can be wind, engine vibrations, music, tire noise, or electrical noise. Coherent noise can significantly be eliminated by using spatial filtering and beamforming methodologies. Non-coherent noise can be reduced by pre-processing the target speech.

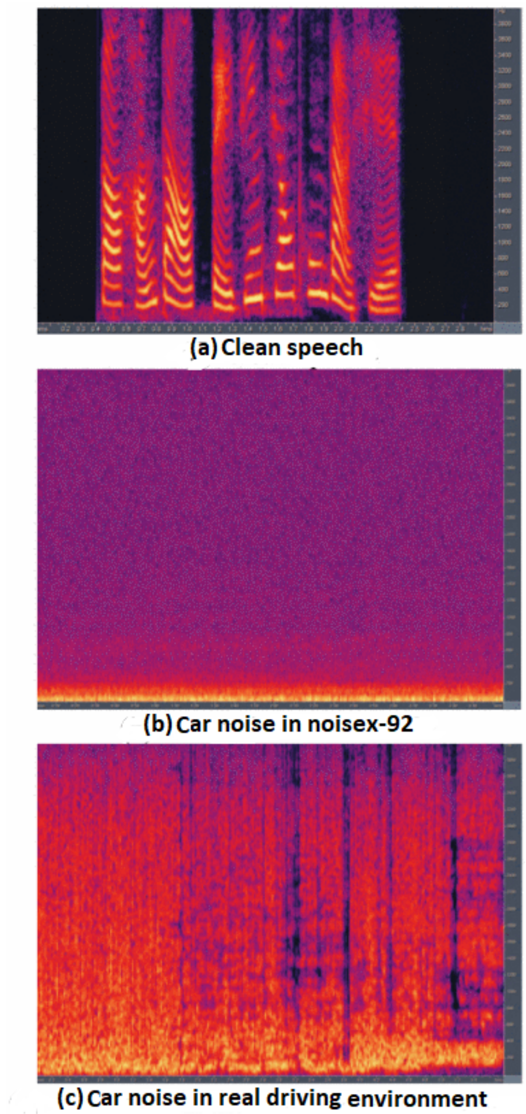

Fig. 8. The spectrogram of clean speech, car noise in noisex-92 and car noise in the real driving environment[1]

In Fig. 8 we can see different spectrograms and their energy distribution in the frequency domain. These figures are generated by car noise collected in a typical driving environment.

It's clear from spectrograms that clear speech has a higher frequency than background noise. Thus, a low-frequency suppression can be added before speech enters the primary channel in GSC so that most of the noise in the environment is suppressed before the beamforming is done. The new improved GSC is as in Fig. 9.

B. Blocking Matrix Improvement by Using Source Position

The purpose of BM is to eliminate target speech from input signals so that the signals that enter adaptive filter contains only noise. In general, the selection of the blocking matrix determines the output effect and complexity of the computations. In general blocking matrix used in traditional GSC is

$$B = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -1 \end{bmatrix}_{M-1 \times M} \tag{15}$$

The design of the BM includes complete knowledge of the direction of arrival of the desired voice signal for microphone signal time alignment.The target speech signal leaks into noise references, resulting in beamformer output canceled for the signal, with an estimating error in the direction of enter and the reflections of signals from artifacts and walls.

For near field array using a traditional BM can cause serious errors in target speech. Moreover, compensating the signal amplitude at the input will produce the overestimation of noise [8]. Hence blocking matrix can be improved by using source position knowledge to obtain balance in signal amplitude. The new blocking matrix is

$$B' = \begin{bmatrix} 1 & -\frac{\sqrt{(r\sin\theta)^2+(r\cos\theta+d)^2}}{r} & 0 & \dots & 0 & 0 \\ 0 & 1 & -\frac{\sqrt{(r\sin\theta)^2+(r\cos\theta+2d)^2}}{\sqrt{(r\sin\theta)^2+(r\cos\theta+d)}} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -\frac{\sqrt{(r\sin\theta)^2+(r\cos\theta+(M-1)d)^2}}{\sqrt{(r\sin\theta)^2+(r\cos\theta+(M-1)d)^2}} \end{bmatrix} \tag{16}$$

Where θ is the angle between the source and reference microphone, r is distance of source from microphone and d is distance between adjacent microphones.

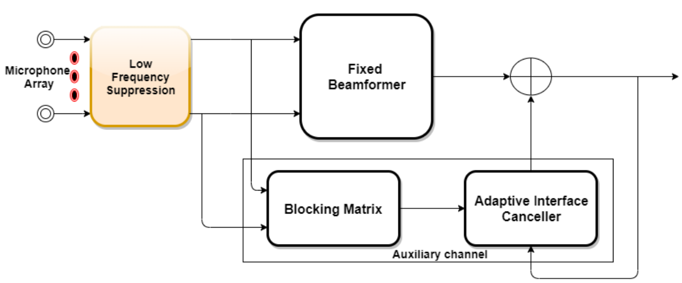

Fig. 9. Improved GSC[1]

C. Improved GSC with Multi-band Spectral Subtraction Algorithm

The beamforming ability of microphone array varies for different frequency bands. Microphone array can effectively suppress the coherent noise but in case of non-coherent noise it must be improved. This ability of microphone can be improved by using Spectral Subtraction Algorithm (SSA).

Spectral subtraction is a method of restoring the power or amplitude of a signal contained in additional noise by removing an approximation of the average noise from the noise spectrum. Thus, by a multi-band SSA is combined in improved GSC to increase the ability to handle non-coherent noise and simplify the computation process [9][10][11].

A spectral subtraction algorithm was proposed in the literature [12] based on the multi-band communicative frequency band divided into various sub-bands N and the spectral subtraction method is separately applied for each sub-band independent of each other [19].

V. Experimental Results

A. Experiment Condition

Four microphone uniform linear array is used in the experiment, with an adjacent distance of 10cm between each microphone. The source i.e. driver is at 60 degrees and 45 cm from the reference microphone. The noisy speech is obtained from a car driving at a speed of 40km/h in a natural street with its windows closed and without passengers.

B. Experiment Result

There have been 60 utterances tested recorded in the driving area. Table II shows the impact of an improved speech on example sentence of the classic GSC , enhanced GSC, enhanced GSC in conjunction with multi-band subtraction in terms of SNR evaluated. By using improved GSC with multi-band SSA the input SNR is almost doubled.

TABLE II
Results of Example Sentence Evaluated[1]

Input SNR(dB)	**Traditional GSC SNR(dB)**	**Improved GSC SNR(dB)**	**GSC with Multi-band SSA SNR(dB)**
5	7.2160	9.4320	11.1238
6	7.7901	9.3821	11.1366
7	8.2354	9.5004	12.3878

Furthermore, Table III shows the corresponding PESQ scores of different algorithms. Perceptual Evaluation of Speech Quality(PESQ) is standardization to measure a degraded voice sample. It returns a score of 4.5 to -0.5, where higher scores indicate better quality. By observing results it is obvious that in terms of denoising efficiency, the proposed algorithm exceeds the conventional one.

For one of the experiment declarations, Fig. 10 shows the reference signal spectrogram, a noisy speech signal of a single microphone, Conventional GSC performance and improved GSC with SSA. One can see the SNR on example speech of the proposed algorithm improved substantially by 3.80 on average dB compared to traditional algorithms. The pitch of speech and its harmonics are more distinctive while the energy from background noise has a sharp decline.

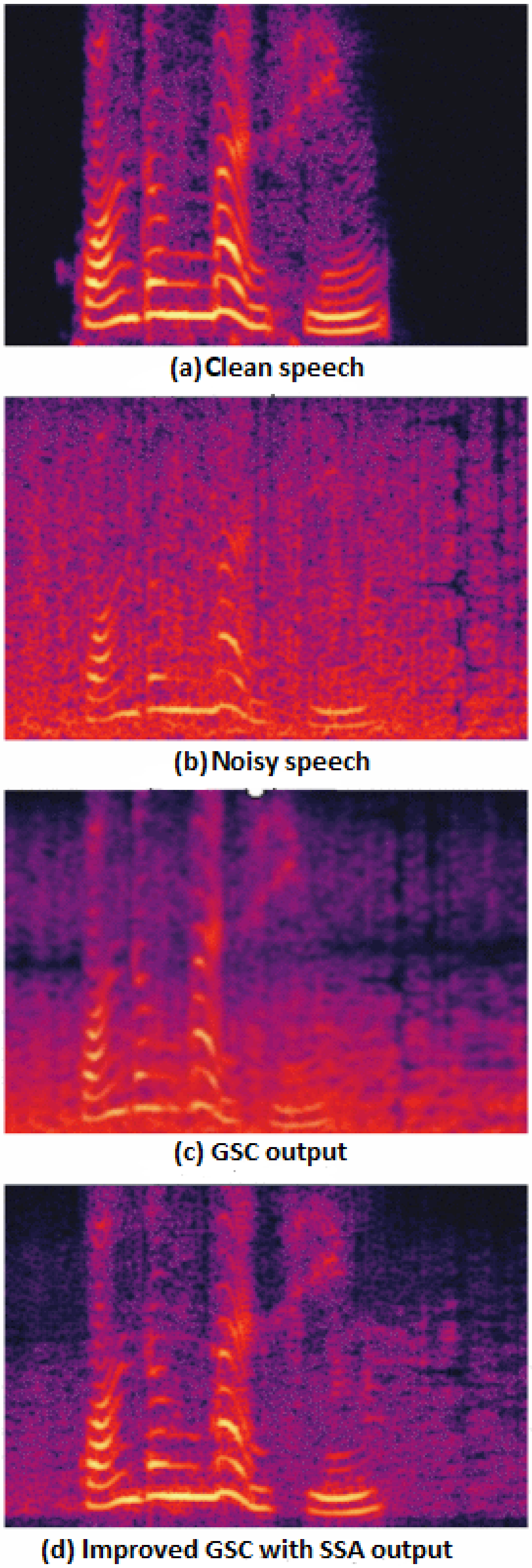

Fig. 10. The spectrogram of the reference signal, noisy speech signal of a single microphone, the output of the traditional GSC and the proposed algorithm[1]

TABLE III
PESQ SCORES OF ALL ALGORITHMS[1]

Input PESQ	**Traditional GSC PESQ**	**Improved GSC PESQ**	**GSC with Multi-band SSA PESQ**
2.0799	2.1172	2.4166	2.6324
2.2176	2.2361	2.5345	2.8176
2.4509	2.5731	2.7725	2.8176

VI. CONCLUSION

MEMS microphone array with multi-channel speech enhancement algorithm based on GSC is proposed in this paper. In pre-processing, low-frequency suppression is used and the BM is improved by considering the wave front curvature of the near-field spherical wave to eliminate the target speech leakage. To reduce the non-coherent noise, a multi-band spectral subtraction algorithm is also adopted. The experimental results show that the proposed algorithm is efficient to replace traditional algorithm used.

REFERENCES

[1] W. Jin, J. Wei and X. Zhong, "Multi-channel speech enhancement in driving environment," 2017 11th International Conference on Signal Processing and Communication Systems (ICSPCS), Gold Coast, QLD, 2017, pp. 1-5, doi: 10.1109/ICSPCS.2017.8270478.

[2] Z. Wang, Q. Zou, Q. Song and J. Tao, "The era of silicon MEMS microphone and look beyond," 2015 Transducers - 2015 18th International Conference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS), Anchorage, AK, 2015, pp. 375-378, doi: 10.1109/TRANSDUCERS.2015.7180939.

[3] Gannot S, Burshtein D,Weinstein E. Signal Enhancement Using Beamforming and Nonstationarity with Applications to Speech [J]. IEEE Transactions on Signal Processing, 2001, 49(8): 1614-1626.

[4] "In-Car Voice Assistant Consumer Adoption Report" 2020. [online] Available at: https://voicebot.ai/wp-content/uploads/2019/01/in-car_voic_assistant_consumer_adoption_report_2019_voicebot.pdf [Accessed 5 June 2020].

[5] "MEMS Microphones, Why Is Digital Best? - CML Microcircuits". [online] Available at: https://www.cmlmicro.com/mems-microphones-why-is-digital-best/ [Accessed 26 June 2020].

[6] Dongxia Wang, Jiachao Zheng et al. A broadband beamforming method based on microphone array for the speech enhancement [A]. International Conference on Signal Processing Systems [C], 2010(2):v2- 363 - v2-366.

[7] Wang Dong-xia and Yin Fuliang. A Nearfield Beamforming Method for microphone Array Based on Speech Enhancement [J]. Journal of Electronics and Information Technology, 2007, 29(1):67-70.

[8] Zheng Y R and Goubran R A. Robust near-field adaptive beamforming with distance discrimination [J]. IEEE Trans. on Speech and Audio Processing, 2004, 12(5): 478-488.

[9] Gannot S and Cohen I. Speech enhancement based on the general transfer function GSC and post filtering [J].IEEE Trans. on Speech and Audio Processing, 2004, 12(6): 561-571.

[10] Hoshuyama O,Sugiyama A,Hirano A Robust Adaptive Beamformer with a Blocking Matrix Using Coefficient-Constrained Adaptive Filters [J]. IEICE Transactions on Fundamentals of Electronics Communications and Computer Sciences,1997:367-370.

[11] R. S. Fazzio, T. Lamers, O. Buccafusca, A. Goel and W. Dauksher, "Design and Performance of Aluminum Nitride Piezoelectric Microphones," TRANSDUCERS 2007 - 2007 International Solid-State Sensors, Actuators and Microsystems Conference, Lyon, 2007, pp. 1255-1258, doi: 10.1109/SENSOR.2007.4300365

[12] Shah, M., Shah, I., Lee, D. and Hur, S., 2019. Design Approaches of MEMS Microphones for Enhanced Performance. Journal of Sensors, 2019, pp.1-26.

[13] Fox C, Vitte G et al. A subband hybrid beamforming for in-car speech enhancement [A]. European Signal Processing Conference [C],Bucharest Romania, 2012: 11-15

[14] J. Hong, S. Han, S. Jeong and M. Hahn, "Adaptive microphone array processing for high-performance speech recognition in car environment," in IEEE Transactions on Consumer Electronics, vol. 57, no. 1, pp. 260-266, February 2011, doi: 10.1109/TCE.2011.5735511.

[15] Z. Chen, Y. Liu, G. Wang, S. Wang and W. Geng, "Multiband Spectral Subtraction Speech Enhancement Algorithm with Phase Spectrum Compensation," 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China, 2019, pp. 2681-2685, doi: 10.1109/IAEAC47372.2019.8997837.

[16] N. Upadhyay and A. Karmakar, "A Perceptually Motivated Multi-Band Spectral Subtraction Algorithm for Enhancement of Degraded Speech," 2012 Third International Conference on Computer and Communication Technology, Allahabad, 2012, pp. 340-345, doi: 10.1109/ICCCT.2012.75.