# In-Car Speech Separation via MVDR

Adilson Chinatto

***Abstract*— In this work, a comparative study about in-car microphone array beamforming is carried out. The main objectives of this work are to show how the number of microphones and the array geometry impact in the task of speech separation via beamforming in in-car environment. These studies lead to suggestions about which topologies are more indicated to be applied in real-world automobile applications, aiming the next-generation autonomous cars. The beamformer considered is the broadband version of MVDR algorithm and three types of arrays, located at the car console central position, are simulated: linear arrays, circular arrays and elliptic arrays.**

***Keywords*— Microphone array processing, speech source separation, broadband beamforming.**

## I. Introduction

Microphone array processing has been an active research theme, finding applications in many fields as in excavation [1], aerodynamic and aero-acoustic testing [2], sports and spectacles [3], health care [4], among many others. One of its main applications lies in speech separation and noise cancellation in the so-called "cocktail-party" problem [5], [6]. Several algorithms and implementations have been developed in the past years addressing to solve the cocktail-party problem, with especial attention to the spatial filtering [7], artificial neural networks (ANN) [8], and to the blind source separation (BSS) [9], [10] techniques.

Although producing quality results, ANN and BSS techniques are often computationally intensive. This prevents their immediate deployment in low cost (unconnected) equipment. On the other hand, spatial filtering techniques are easier to implement and can present satisfactory results in problems related to speech-source separation and direction of arrival (DOA) estimation. When used as a pre-processing module in the speech processing chain, spatial filtering can lower the complexity and convergence time of ANN and BSS algorithms [11], [12]. In other words, beamforming via spatial filtering can give a cleaner pre-processed speech for later enhancement by ANN or BSS.

Microphone array beamforming is not a recent research subject [13]. This technique is derived from the previous works related to the antenna arrays. However, microphone array beamforming has gained especial attention in the recent years with the growing applications in consumer appliances, whose examples range from the next-generation smart-houses to the autonomous cars. Especially in this latter application, voice discrimination and recognition play an important role in executing crucial voice commands that can directly impact in the comfort and safety of the passengers.

Several research groups have been carrying out in-car beamforming studies, mainly focused in the driver speech enhancement. However, few works have been presented considering aspects as array geometry and position inside the car and their relation with the quality of the speech separation via beamforming. This work points out considerations about the best array positioning inside the car and the performance of speech separation via beamforming is compared in terms of array geometries and its number of elements.

Here, the algorithm chosen to spatial filtering implementation is the well-known minimum variance distortionless response (MVDR) in its broadband formulation. MVDR was initially proposed in the 1980's [14] and was quickly adopted in applications related to electromagnetic waves. Its use in microphone array beamforming was proposed in the 1990's [15] and even recently it has been an active research theme [16]. For its implementation, three types of geometries are considered: uniform linear array (ULA), uniform circular array (UCA) and uniform elliptical array (UEA).

The paper is organized as follows: in section II, the problem about the array position inside the car is addressed and some considerations basis the choice considered in the rest of the paper; Section III brings the mathematical fundamentals of broadband MVDR algorithm devoted to spatial speech separation; Section IV shows the simulation results and brings discussion about the impact of array geometry and number of elements; finally, section V presents the work conclusions.

## II. Considerations on In-Car Array Positioning

Positioning a microphone array inside a car needs to obey some important requisites related to the response quality, installation costs and housekeeping easiness. Several works dealing with in-car speech separation and enhancement adopt two main microphones dispositions: microphones spread in the whole car interior, especially on the roof and microphones circling the conductor. Although presenting some interesting results, these types of disposition present some drawbacks that can difficult its adoption in commercial systems. First of all, spreading the microphones throughout the car interior make the receivers experiment different levels of noise, produced by the engine and friction. This can lead to degradation in the beamforming response. Moreover, considering that the objective is to enhance the speech of the conductor and passengers, microphones located in the car roof are not positioned in the region in which the speech would be stronger, degrading the results. On the other hand, microphones circling the conductor seems to be a good solution to enhance the speech of this person, but barely will produce good results in speech enhancement of the others car occupants.

Adilson Chinatto: Venturus Instituto de Tecnologia, Campinas-SP, Brazil, e-mail: adilson.chinatto.xx@venturus.org.br; School of Electric and Computation Engineering, University of Campinas (UNICAMP), Campinas-SP, Brazil, e-mail: chinatto@gmail.com; Espectro Ltda., Campinas-SP, Brazil, e-mail: chinatto@espectro-eng.com.br

Another drawback in the spreading the microphones throughout the car interior is the cabling needs. The multi-channel speech processing is generally performed in a single board computer. In this way, it is necessary to connect all the microphones to this unit. If the microphones are spread in the car interior their usage will demand long signal and power cables. These long cables can be a source of interference in the signal processing chain and increase the total system cost.

Generally, car on-board computers are located inside the car front panel. Moreover, all car occupants are positioned in a way that the speech is projected to the front, towards the front panel. In order to keep the microphones closer, in a neutral position in relation to the speech projections of the car occupants, and trying to reduce the cabling, the best position to place the microphone array is in the center of the car panel. Figure 1 shows schematically the array position inside a car adopted in this work.

## III. Problem Formulation and Algorithms

In this work, an $M$-element array formed by microphones located at positions $(x_{a_m}, y_{a_m}, z_{a_m})$, $m = 0, 1, \cdots, M-1$ is considered. This array is impinged by spherical sound waves created by $L$ speech sources located at positions $(x_{s_\ell}, y_{s_\ell}, z_{s_\ell})$, $\ell = 0, 1, \cdots, L-1$. The output of the $m$-th microphone is given by

$$y_m(t) = \sum_{\ell=0}^{L-1} s_\ell(t - \tau_{m\ell}) + \nu_m(t), \tag{1}$$

where $s_\ell(t)$ is the $\ell^{th}$ speech source and $\tau_{m\ell}$ is the time delay between the $\ell^{th}$ source speech and the $m^{th}$ microphone, given by

$$\tau_{m\ell} = \frac{\|(x_{a_m}, y_{a_m}, z_{a_m}) - (x_{s_\ell}, y_{s_\ell}, z_{s_\ell})\|_2}{v}. \tag{2}$$

In (2), $v$ is the speed of the sound in the air. Finally, the microphone output $y_m(t)$ is corrupted by the noise $\nu_m(t)$, considered to be Gaussian.

In a such system, the signals at the microphones can be weighed and combined to enhance signals received under desired directions while mitigating signals received under others undesired directions. This beamforming process can be implemented via several well-known techniques as delay-and-sum or generalized sidelobes cancellers, for instance. In this work, the broadband version of the MVDR is considered.

### A. Narrowband MVDR Formulation

The MVDR algorithm formulation is originally attributed to Capon [17] and is widely applied in antenna arrays. Be the system of equations $\mathbf{y}(n) = \mathbf{A}(\phi, \theta)\mathbf{s}(n) + \boldsymbol{\nu}(n)$ in which $\mathbf{s} = [s_0(n)\ s_1(n)\ \cdots\ s_{L-1}(n)]^T$ is a vector formed by the $n^{th}$ sample of the $L$ incident signals, $\boldsymbol{\nu}$ is a noise vector and $\mathbf{A}(\phi, \theta) = [\mathbf{a}(\phi_0, \theta_0)\ \mathbf{a}(\phi_1, \theta_1)\ \cdots\ \mathbf{a}(\phi_{L-1}, \theta_{L-1})]$ is a mixture matrix that contains the steering vectors for each incident signal related to the array. For the sake of simplicity, considering a uniformly linear array (ULA), the contribution from $\theta$ is null and the steering vector related to the $\ell^{th}$ signal is given by an $M$-element vector $\mathbf{a}(\phi_\ell)$, whose $m^{th}$ component is defined as $e^{-j2\pi md\frac{c}{f_c}\sin\phi_\ell}$, $m = 0, 1, \cdots, M-1$. In this formulation, $d$ is the distance between array elements, $c$ is the speed of propagation (correspondent to the speed of light in electromagnetic systems) and $f_c$ is the carrier frequency, considered to be much bigger than the signal bandwidth. By convention, $\phi_\ell = 0°$ is normal to the line formed by the array elements and the element correspondent to $m = 0$ is the reference one.

The problem here is to determine the set of weight coefficients $\mathbf{w}$ to be applied to the $M$-sensors output $\mathbf{y}$ in order to recover a desired signal $s_D(n)$ from the *a priori* knowledge of $\phi_D$, while mitigating signals contributions from others directions. This is the solution of

$$\min_{\mathbf{w}} \mathbf{w}^H\mathbf{R}\mathbf{w} \text{ s. t. } \mathbf{w}^H\mathbf{a}(\theta_D) = 1, \tag{3}$$

where $\mathbf{R} = E\{\mathbf{y}\mathbf{y}^H\}$, $E\{.\}$ is the expectation operator and $\mathbf{a}(\theta_D)$ is the steering vector of the desired angle of arrival $\theta_D$. The Capon solution of (3) is given by

$$\mathbf{w}_C = \frac{\mathbf{R}^{-1}\mathbf{a}_D}{\mathbf{a}_D^H\mathbf{R}^{-1}\mathbf{a}_D}, \tag{4}$$

where $\mathbf{a}(\theta_D)$ was substituted by $\mathbf{a}_D$ for clarity reasons. Evaluating $\hat{s}_D(n) = \mathbf{w}_C^H\mathbf{y}(n)$ recovers $s_D(n)$ with unit gain while mitigating other signals that impinge the array under different steering vectors.

### B. Broadband MVDR Formulation

Considering that the observables $y_m(t)$ in (1) are sampled at times $t = nT_S$, where $T_S$ is the sampling period and $n \in \mathbb{N}$, then, after $N$ samples, the outputs $y_m(nT_s)$ can be translated into the frequency domain via $N$-length discrete Fourier transform ($N$-DFT) leading to be represented by a complex $M \times N$ matrix $\mathbf{Y}(\omega)$, where $\omega = 2\pi k/N$, $k = 0, 1, \cdots, N-1$. Considering that the original temporal $L$ incident signals can be represented in the frequency domain by a $L \times N$ matrix $\mathbf{S}(\omega)$, for each frequency bin $\omega$, the problem can be stated as

$$\mathbf{Y}(\omega) = \mathbf{A}(\tau, \omega)\mathbf{S}(\omega) + \mathbf{N}(\omega), \tag{5}$$

where $\mathbf{A}(\tau, \omega) \in \mathbb{C}^{M \times L}$ is the array of steering vectors formed by the time-delay of arrival (TDOA) related to each microphone at frequency bin $\omega$, given by $\mathbf{A}(\tau, \omega) = [\mathbf{a}(\tau_0, \omega)\ \mathbf{a}(\tau_1, \omega) \cdots\ \mathbf{a}(\tau_{L-1}, \omega)]$, with

$$\mathbf{a}(\tau_\ell, \omega) = \begin{bmatrix} 1 \\ e^{-j\frac{\omega}{N}\tau_{1,\ell}} \\ e^{-j2\frac{\omega}{N}\tau_{2,\ell}} \\ \vdots \\ e^{-j(M-1)\frac{\omega}{N}\tau_{M-1,\ell}} \end{bmatrix}. \tag{6}$$

It is easy to notice that (6) is similar to (**??**), however, the dependence of $\omega$ makes the problem to be broadband. Due to this frequency dependence, MVDR should be evaluated for every DFT frequency bin $\omega$, resulting in $N$ weighing vectors $\mathbf{w}_{C,\omega}$. Finally, to avoid division by zero, a little constant $r$ acting as a regularization parameter can be added, resulting in

$$\mathbf{w}_{C,\omega} = \frac{\mathbf{R}_\omega^{-1}\mathbf{a}(\tau_D, \omega)}{\mathbf{a}(\tau_D, \omega)^H\mathbf{R}_\omega^{-1}\mathbf{a}(\tau_D, \omega) + r}, \tag{7}$$

where $\mathbf{w}_{C,\omega}$ and $\mathbf{R}_\omega$ are the Capon solution and the correlation matrix related to the $k^{th}$ frequency bin respectively, and $\mathbf{a}(\tau_D, \omega)$ is the steering vector related to the desired TDOA $\tau_D$ at the frequency bin $\omega$, given by (6).

In the frequency domain, the desired signal is estimated by evaluating

$$\hat{\mathbf{S}}_D(\omega) = \mathbf{w}_{C,\omega}^H \mathbf{Y}(\omega), \tag{8}$$

for each $\omega = 2\pi k/N$, $k = 0, 1, \cdots, N-1$. Finally, the temporal broadband MVDR estimation $\hat{s}_D(n)$ is given by the inverse of the $N$-point DFT of $\hat{\mathbf{S}}_D(\omega)$.

## IV. Numerical Results on Simulated Data

In this section, the experiment is described and the simulations results are presented and analysed.

### A. Array Geometries and Speech Sources Locations

The in-car speech separation using broadband MVDR beamformer was simulated considering three array geometries: UCA, ULA and UEA. Each array was considered to be centered in the car panel, as schematically depicted in Figure 1, which is the best position of an electronic device inside a car, as mentioned in Section II. The microphone positions for each array geometry are given by

$$(x_{a_m}, y_{a_m}, z_{a_m}) =$$

$$\text{ULA} := \left(\frac{d(M-1+2m)}{2(M-1)}, 0, 0\right) \tag{9a}$$

$$\text{UCA} := \frac{d}{2}\left(\cos\left(\frac{2\pi m}{M}\right), \sin\left(\frac{2\pi m}{M}\right), 0\right) \tag{9b}$$

$$\text{UEA} := D\left(\cos\left(\frac{2\pi m}{M}\right), \alpha \sin\left(\frac{2\pi m}{M}\right), 0\right), \tag{9c}$$

where $m = 0, 1, \cdots, M-1$, $d$ is a real positive scalar equivalent to the radius in the UCA and the array length in the ULA, $D$ is the length of the ellipse major axis in the UEA geometry and $\alpha$ is a real positive scalar correspondent to the relation between the lengths of the ellipse's semi-minor and semi-major axes in the same geometry.

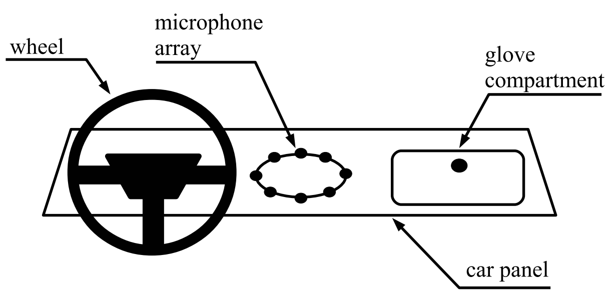


Fig. 1. Car panel representation showing the array position considered. The array geometry and number of elements are only illustrative.

Two speeches were considered to be emitted by persons sit at three possible positions inside the car: at the driver position, at the front-right passenger position, or at the back-seat passenger position, as schematically shown in Figure 2. Although there are three possible positions, it is considered that only two car occupants speak at the same time. This approach is realistic as we are considering that the car driver is giving a voice command at the same time that the others two car occupants talk, speaking one of its time. The in-car position of each car occupant is given by

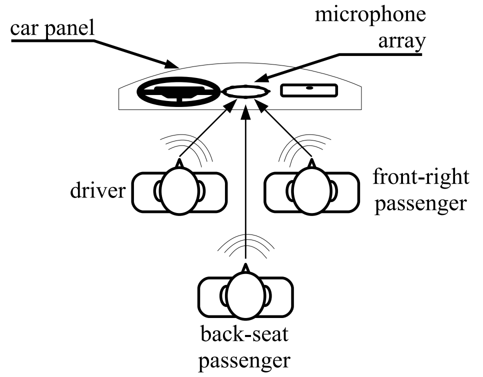


Fig. 2. Speech sources positions inside a car.

- driver: $(x_s, y_s, z_s) = (+0.350, 0.000, +0.606)$ m
- front-right: $(x_s, y_s, z_s) = (-0.350, 0.000, +0.606)$ m
- back-seat: $(x_s, y_s, z_s) = (0.000, 0.000, +1.100)$ m

### B. Simulations Parameters

Two real prerecorded 1365.3 ms duration voices, one male and one female, were used in the experiments. The voices were recorded monoaurally, under a sampling rate $f_S = 48$ kHz, and inside a car in movement. In this way, 65536 samples were considered as output of each microphone. The male voice corresponds to a short driver car command and the female voice corresponds to a extract of a chit-chat.

In all simulations, the window length $N$ was fixed in 1024, what means that the $N$-DFT and -IDFT were the size 1024. For the three array types, the number of microphones $M$ was varied from 6 to 16. For ULA and UCA, $d$ was changed from 2 cm to 30 cm, what implies in very dense arrays to somewhat sparse ones. It is important to notice that all these array geometries can be implemented in the practice, considering that nowadays there are surface mount technology (SMT) microphones with dimensions as smaller as $2.50\,\text{mm} \times 1.60\,\text{mm}$ [19]. Finally, for UEA, $D$ was fixed in 30 cm and $\alpha$ was changed from 0 to 1.

### C. Practical aspects aiming implementation

The weighting vectors $\mathbf{w}_{C,\omega}$ are evaluated at each $N$ samples received at the outputs of the $M$ microphones. As initially pointed out in [18], some minor modifications can be performed in (7) to enable easier and more robust practical implementation. One of these modifications is related to the frequency bins to be used in the MVDR calculation. Theoretically, using $N$ frequency bins implies in $N$ partial MVDR evaluations, one for each bin. However, information in the speech, when treated at higher sampling rates, can be considered to be concentrated only in a narrow frequency band $B$ most of the time. So, MVDR evaluation on frequency bins beyond the band $B$ results in negligible contribution in the speech separation. In the simulations, $f_S = 48\,\text{kHz}$ and knowing that the information in the speech is almost all concentrated bellow 3.6 kHz, the MVDR was evaluated for frequency bins under $B = 4\,\text{kHz}$.

Another modification is related to the autocorrelation $\mathbf{R}_\omega$ evaluation. In practical systems it is impossible to exactly evaluate $\mathbf{R}_\omega$ as it would demand infinite samples. Alternatively, $\mathbf{R}_\omega$ can be approximated by using the sample covariance matrix instead, given by

$$\hat{\mathbf{R}}_\omega(\eta) = \frac{F-1}{F}\hat{\mathbf{R}}_\omega(\eta-1) + \frac{1}{F}\mathbf{Y}_\omega(\eta)\mathbf{Y}_\omega(\eta)^H, \tag{10}$$

where $\mathbf{Y}_\omega(\eta)$ corresponds to the $\mathbf{Y}(\omega)$ at the $\eta^{th}$ frame, $\hat{\mathbf{R}}_\omega(\eta)$ is the sampled covariance matrix of the frequency bin $\omega$ at the $\eta^{th}$ frame and $F$ is a weighting constant that acts as a forgetting factor. In the simulations $F$ was empirically set to 256.

Finally, to cover cases in which the voices suffers with quick variations, the broadband MVDR was evaluated for frames compound by $N$ samples, as previously defined, but the difference in samples between the frame $\eta$ and the frame $\eta+1$ was set in $L = 16$. So, the weighting vectors were updated at each $L/f_S = 3.333\,\mu\text{s}$ against the $N/f_S = 21.333\,\text{ms}$, which would be the update rate if the overlapping would not be used.

### *D. Results*

The results presented in this section consider the signal-to-interference-plus noise (SINR), evaluated following [20], which considers that the estimated signal can be represented as

$$\hat{s}_L = s_L + e_{L,inter} + e_{L,noise} + e_{L,artif}, \tag{11}$$

where $s_L$ is the original $L^{th}$ source, $\hat{s}_L$ is its estimate, $e_{L,inter}$ is the estimates deformation due to the interferer, $e_{L,noise}$ is the deformation due to the noise and $e_{L,artif}$ is the estimates deformations due to imprecision in the estimating processes. So, the SINR is given by

$$\begin{aligned}\text{SINR} &= 20\log\frac{\|s_L\|}{\|e_{L,inter} + e_{L,noise} + e_{L,artif}\|}\\ &= 20\log\frac{\|s_L\|}{\|\hat{s}_L - s_L\|}.\end{aligned} \tag{12}$$

Figures 3 and 4 show the simulation results for UCA and ULA with several numbers of microphones and a set of array dimensions. These figures consider the case of driver and front-right passenger speech separation. As expected, separation increases monotonically with the number of microphones and with the UCA diameter or with the ULA length. The same array geometries were used in the task of speech separation between the driver and back-seat passenger. These results are shown in the figures 5 and 6. Again, the results are similar between UCA and ULA.

It is interesting to notice that the separation provided by the MVDR is stronger when dealing with driver and back-seat passenger. In fact, there is a gain about 2 dB when compared with the situation in which the separation is performed between driver and front-right passenger speeches. At the

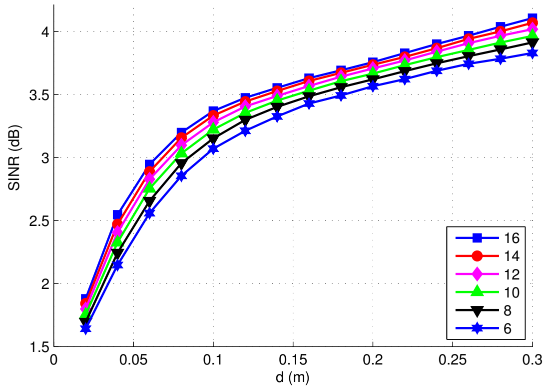


Fig. 3. SINR for UCA: front-side passenger.

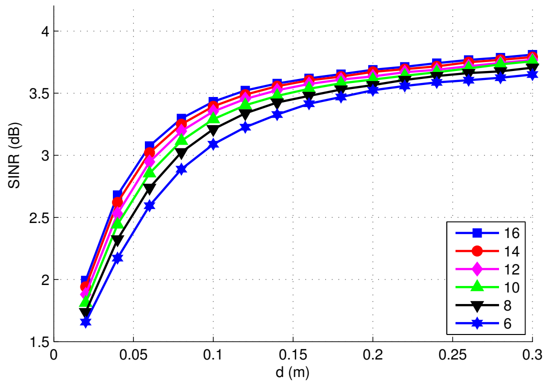


Fig. 4. SINR for ULA: front-side passenger.

same time, the number of receivers has little influence in the speech separation. No matter if the scenario corresponds to the driver and front-right passenger or to the driver and back-seat passenger separation, changing the number of microphones from 6 to 16 impacts in a gain of at most 1 dB. To investigate the algorithm performance behavior in the transition from ULA and UCA, a set of simulations was performed using EUA with fixed semi-minor axis $D = 0.15\,\text{m}$ and changing the relation between the two ellipse axis 0 (ULA) to 1 (UCA). As can be noticed, the changes in the performance can be considered negligible. Simulation results show that there is no significant speech separation for $d > 0.1\,\text{m}$ in the case of driver – left-front passenger. For the case driver – back-seat passenger, the same occurs for $d > 0.2\,\text{m}$, although for larger arrays the algorithm performs even better. Moreover, the geometry analysed presents little influence in the algorithm performance. This leads to infer that in this in-car application, using an array with $d = 0.15\,\text{m}$, linear, circular or even elliptic is the best compromise between performance and cost.

## V. Conclusions

In this paper, the ideal position of a microphone array inside a car aiming to receive voice commands was debated. From this debate, the broadband MVDR beamformer was used in the task of in-car speech separation. It was inferred that the best position to place a microphone array inside a car is the center of the car console since in this position it is easy to do all the cabling and the array would be able to receive

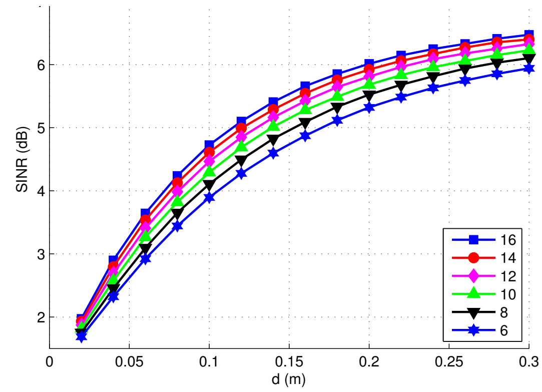


Fig. 5. SINR for UCA: back-sit passenger.

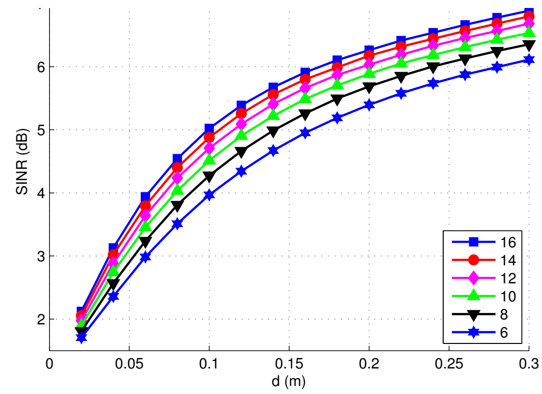


Fig. 6. SINR for ULA: back-sit passenger.

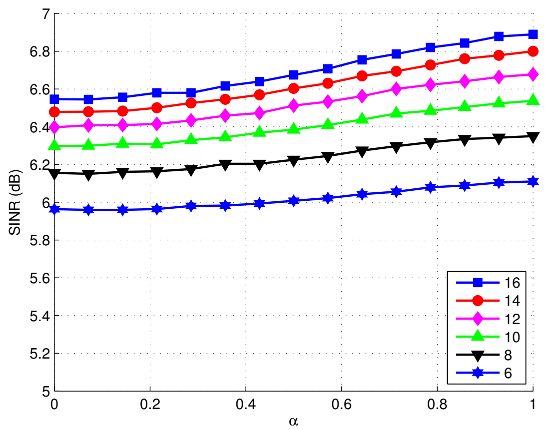


Fig. 7. SINR for UEA: back-sit passenger.

and process all the voices. Moreover, the simulation results of the broadband MVDR beamformer show that arrays with a horizontal dimension from 0.10 m to 0.20 m and 6 to 16 elements can provide good SINR in-car speech source separation. Interestingly, at the problem addressed, the array geometry presents little impact in the speech separation performed via broadband MVDR beamformer. This observation is important as it allows to keep the array geometry as simple as possible, reducing the occupied area, the computational cost and the implementation costs.

## ACKNOWLEDGEMENTS

The author would like to thank the VENTURUS – CENTRO DE INOVAÇÃO TECNOLÓGICA for funding, support, and motivate this work.

## REFERENCES

[1] Tianlei Wang, Jiuwen Cao, Jianzhong Wang, Xiaoping Lai, and Zhiping Lin, "Excavation devices classification using enhanced acoustics by MVDR beamforming with a cross microphone array," *in Proceedings of the IEEE International Confernce on Digital Signal Processing (DSP)*, pp. 388-392, 2016.

[2] T. Ito, H. Ura, and Y. Yokokawa, "Arrayed microphone measurement in low speed wind tunnel," *in Proceedings of the 2005 International Congress on Instrumentation in Aerospace Simulation Facilities (ICI-ASF)*, pp. 40-46, 2005.

[3] Nikolaos Stefanakis and Athanisios Mouchraris, "Capturing and reproduction of a crowded sound scene using a circular microphone array," *in Proceedings of the 24th European Signal Processing Conference (EUSIPCO)*, pp. 1673-1677, 2016.

[4] Yun Li, K. C. Ho, Mihail Popescu, and Marjorie Skubic, "A theoretical study on placement of microphone array for improving the localization accurary of a fall," *in Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4523-4526, 2014.

[5] Colin E Cherry "Some Experiments on the Recognition of Speech, with One and with Two Ears," *The Journal of the Acoustical Society of America*, vol. 25, issue 5, pp. 975-79, 1953.

[6] J. Barrere and G. Chabriel, "A compact sensor array for blind separation of sources," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 49, issue 5, pp. 565-574, 2002.

[7] S. Doclo and M. Moonen, "Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics," *IEEE Transactions on Signal Processing*, vol. 51, issue 10, pp. 2511-2526, 2003.

[8] Aditya Arie Nugraha, Antoine Liutkus, and Emmanuel Vincent, "Multichannel audio source separation with deep neural networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, issue 9, pp. 1652-1664, 2016.

[9] A. Bell and Terrence Sejnowski. "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, issue 6, pp. 1129-1159, 1995.

[10] Taesu Kim, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, issue 1, pp. 70-79, 2007.

[11] S Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," *IEEE Transaction on Speech and Audio Processing*, vol. 11, issue 2, pp. 109-116, 2003.

[12] L. Parra and C. Fancourt, "An adaptive beamforming perspective on convolutive blind source separation," *in G. M. Davis (ed.) Noise Reduction in Speech Applications*, CRC Press, 2002.

[13] J. L. Flanagan, "Beamwidth and useable bandwidth of delay-steered microphone arrays," *AT&T Technical Journal*, vol. 64, issue 4, pp. 983-995, 1985.

[14] Tie-Jun Shan and T. Kailath, "Adaptive beamforming for coherent signals and interference," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, issue 3, pp. 527-536, 1985.

[15] S. Gazor and Y. Grenier, "Criteria for positioning of sensors for a microphone array," *IEEE Transactions on Speech and Audio Processing*, vol. 3, issue 4, pp. 294-303, 1995.

[16] Saurav R. Tuladhar and John R. Buck, "Unit circle MVDR beamformer," *in Procedings of the 2015 IEEE International Conference on Acoutics, Speech and Signal Processing (ICASSP)*, pp. 2484-2488, 2015.

[17] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of IEEE*, vol. 57, issue 8, pp.1408-1419, 1969.

[18] M. E. Lockwood, D. L. Jones, C. R. Lansing, W. D. O'Brien, B. C. Wheeler, and A. S. Feng, "Effect of multiple nonstationary sources on MVDR beamformers," Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, pp. 730-734 , 2003.

[19] Cirrus Logic, CS7331P Series, available online in *https://www.cirrus.com/products/cs7331/*.

[20] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation", *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, issue 4, pp. 1462-1469, 2006.