INVESTIGATION OF EFFECTIVENESS OF MICROPHONE ARRAYS FOR IN CAR USE BASED ON SOUND FIELD SIMULATION.

V. Galanenko, A.Kalyuzhny, A. Kovtonyuk

Scientific-Production Enterprise “DELTA”
27, Acad. Krymskogo Str., Kiev, 252142, Ukraine
aleko@delta-net.kiev.ua

ABSTRACT

The effectiveness of small microphone arrays for in car use is investigated. These arrays are designed for speech enhancing at noise background within the cabin of a moving car. Speech and noise simulation based on designed mathematical algorithm of sound field modeling within a car cavity is applied for predicting effectiveness of the spatial-time processing algorithms. The mathematical model takes into account complicated cabin geometry, the difference of its sizes in wavelength scale at low and high frequencies, frequency dependent sound absorption of the cabin surfaces and distributed noise sources with their cross correlation.Theoretical estimates of the microphone arrays effectiveness and output simulated signals (for subjective estimation) are presented.

1. INTRODUCTION

One of the most promising directions of modern acoustic engineering is the wide application of the hands-free interface for systems of cellular telephony (CT) and speech control in cars and other vehicles. The peculiarity of the hands-free systems if compared to traditional CT is the relatively far distance between the microphone and the source of a voice call. Furthermore the car environment is saturated by many kinds of interfering noises. Speech contamination by noises allays the effectiveness of the speech recognition and communication systems. Additional speech enhancement could improve these systems quality. So it is necessary to use more effective methods of a voice call selection at noise background in car hands-free systems as compared to usual ones.

The promising direction to solve this problem is to use the microphone arrays (MA) which show more ample capabilities for selection of a voice call at noise background comparatively to a single microphone. Different versions of the MA application in rooms are considered for example in [1 - 4]. The additional advantages we can expect from MA usage are due to both distinction of speech and noise spectrums and distinction of their sound fields’ spatial structure. Conjoint exploitation of these distinctions is the essence of spatial-time processing. In this case the speech enhancement system includes an additional block of spatial-time filtration of a speech sound field at noise background. Mentioned distinctions of speech and noise sound fields depend on car cavity geometry, acoustical properties of its surfaces, structure and disposition of speech and noise sources. So we need reliable mathematical models of speech and noise sound fields to investigate the MA capability.

Previous consideration determines the paper structure. In the next section the optimization of microphone arrays is discussed. In the third section we consider in brief the problem of modeling and simulation of sound fields within a car cavity. In the last section some results of numerical computations are presented and discussed.

2. MICROPHONE ARRAY OPTIMIZATION

According to its working conditions the microphone array for in car use must be small and consists of a small number of elements. Small linear array is expected to be most convenient. It could be disposed for example at the top of device board on condition of respective vibroisolation. Let the MA elements are disposed at P arbitrary points with the position vectors $\mathbf{r}_p, p = 1,...,P$. The desired signal is a sequence of values of the speech sound field at array aperture (including both the direct wave and reflections). The interference consists of acoustical noises appeared in a moving car (accounting for all the reflections) and noncorrelated noises in electrical circuits.

As the speech processing involves usually the signal segmentation to short frames (of 10 - 20 ms) we can consider speech signal and noise like stationary ones within the frames. The real speech nonstationarity can be entertained by the trimming of the processing algorithm parameters from one frame to another. On additional condition $L/c << T$ (where L is maximal aperture size, T is the frame length), which is usually true for car MA, it is possible to subdivide spatial-time processing algorithm into spatial processing (dependent on frequency) and further filtration. In this case spatial signal processing of MA outputs follows to such an algorithm:

$$u(\omega) = \sum_{p=1}^{P} L^*(\omega, \mathbf{r}_p) U(\omega, \mathbf{r}_p) = \mathbf{L}^*(\omega)\mathbf{U}(\omega) \qquad (1)$$

where $\mathbf{U} \equiv U(\omega, \mathbf{r}_p)$ are the Fourier transforms of array inputs and $\mathbf{L} \equiv L(\omega, \mathbf{r}_p)$ is the amplitude-phase distribution (APD) of the microphone sensitivity at the MA aperture. The next step of signal processing consists in filtering of signal (1) to select the transmitted voice call spectrum $F(\omega) = u(\omega)H(\omega)$.

Well known full optimization (i.-e. optimal choice of $\mathbf{L}(\omega), H(\omega)$) based on likelihood ratio is not always possible. In this context the partial optimization (when either of the two mentioned functions is optimized following to some criterion and another is chosen basing on “reasonable” consideration) is practically interesting. It leads to lot of quasioptimal algorithms. Let consider $F(\omega)$ as an estimate of speech spectrum. Then minimization of the root-mean-square error of estimation leads to the following:

$$H(\omega) = \frac{g_S(\omega) D_S(\omega)}{g_S(\omega)\left|D_S^2(\omega)\right| + g_N^A(\omega)} \tag{2}$$

Here: $g_S(\omega)$ is statistically averaged speech spectrum, $g_N(\omega)$ is noise spectrum at the array output (with regard to accepted algorithm of spatial processing), $D_S(\omega)$ depends on both spatial processing algorithm and wave propagation law defined by Green's function $G(\omega, \mathbf{r})$:

$$D_S(\omega) = \mathbf{L}^*(\omega)\mathbf{G}(\omega)$$

where $\mathbf{G}(\omega) \equiv G(\omega, \mathbf{r}_p),\ p = 1,...,P$

$$g_N^A = g_n(\omega)\left|\mathbf{L}(\omega)\right|^2 + \mathbf{L}^*(\omega)\mathbf{K}_N(\omega)\mathbf{L}(\omega)$$

where $\mathbf{K}_N(\omega) \equiv K_N(\omega, \mathbf{r}_{p1}, \mathbf{r}_{p2})$ is the noise correlation matrix at array aperture, $g_n(\omega)$ is the electrical noise spectrum. The root-mean-square error of filtration corresponding to such a choice of frequency response $H(\omega)$ is equal to $\varepsilon(\omega) = g_S(\omega)/(1+\mu(\omega))$ where

$$\mu(\omega) = \frac{g_S(\omega)\left|\mathbf{L}(\omega)^*\mathbf{G}(\omega)\right|^2}{\mathbf{L}(\omega)^*(\mathbf{K}_N(\omega) + g_n(\omega)\mathbf{I})\mathbf{L}(\omega)} \tag{3}$$

is the signal to noise ratio (SNR) at spatial processing system output. These expressions allow to estimate the filtration error for arbitrary accepted spatial processing algorithm. By maximizing the value $\mu(\omega)$ we obtain optimal APD as the solution of the following equation (full optimization):

$$(\mathbf{K}_N(\omega) + g_n(\omega)\mathbf{I})\mathbf{L}(\omega) = \mathbf{G}(\omega) \tag{4}$$

Most traditional quasioptimal algorithm (known as beamforming technique) consists in MA phasing to match the APD with a plane incident wave.

$$\mathbf{L}(\omega) = \mathbf{E}^*(\omega) \equiv \exp(-i\mathbf{k}_0\mathbf{r}_p), \quad p = 1,..P \tag{5}$$

Another well-known approach is to apply adaptive procedures. As the speech and noise sources' structure is practically stable in car cavity and their levels and spectrums vary slowly in time, the adaptive algorithms are expected to be effective. One of them is the Capon's algorithm [5] (the adaptive version of one quasioptimal algorithm). Following to this algorithm the APD is the solution of the equation (4) with $\mathbf{E}(\omega)$ in right hand side and with matrix $\mathbf{K}_N(\omega)$ replaced by its estimate $\hat{\mathbf{K}}_N(\omega)$ (obtained during the adaptation process). As the information on speech wave field structure is adjusted, the Capon's algorithm can be refined and advanced to the adaptive version of the optimal algorithm.

The consideration of the MA optimization problem points to the importance of sound field structure concept. We believe that simplified models of speech and noise sound fields (unique plane wave for speech field and isotropic noise field for example) are not satisfactory in the considered case of the car cavity. Mathematical models accepted in this work for computing are posed below in brief.

3. MATHEMATICAL MODEL OF SOUND FIELDS WITHIN CAR CAVITY

Reasoning from general consideration the essential feature of the sound field within a car cavity is its intricate spatial structure, which includes both the direct wave and the numerous reflections coming with some delays. At that, reflected signals are not precise counterparts of initial signal. There are two reasons on which their spectra differ: frequency dependence of reflectance of materials covering surfaces and distinction of sound diffraction at different frequencies. At low frequencies a car cavity should be treated as a room of small sizes in wavelength scale. At high frequencies we have opposite situation, which allow geometrical-optical (ray) approach. As speech is wide band signal, disparate algorithms of sound simulation need to be applied for low and high frequencies. Therefore the known computational methods and the respective software used in room acoustics [6] are not convenient.

Proposed mathematical algorithm of sound field simulation is based on Green's function computing. Initial frequency range is subdivided into two parts (subdivision frequency is chosen from the range 500-1000 Hz). Green's function, which describes the signal transmission from a source to an element of MA, is computed for each frequency of both subbands. When computing Green's function for low frequencies we applied numerical-analytical solving the wave equation by means of technique proposed in [7]. In accord to the developed method the solution is represented by the expansion in eigenfunctions of a boundary problem along coordinate axis which is perpendicular to the cabin side walls. The series coefficients are the functions of two other Cartesian coordinates and have to be found numerically. For this expansion validity it is necessary the side walls to be treated as parallel and the cabin cavity have to be approximated by rectangular prism with the side walls as polygonal bases. An example of such an approximation is shown at Fig. 1.

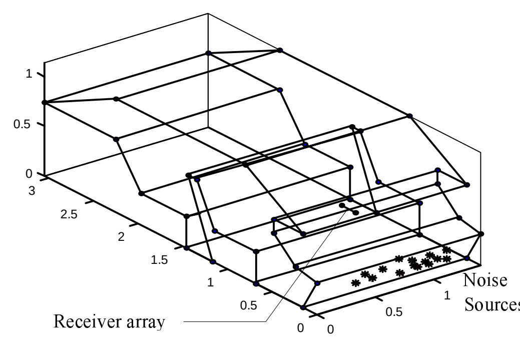

Fig. 1. Car cavity is approximated by a rectangular prism.

Additional assumption is that the sound absorption of side walls can be described for each frequency by areamean absorption coefficient. This assumption is justified for sufficiently low frequencies. It is assumed, besides that, the sound absorption of the side prism faces does not depend on coordinate along prism height (but may be different for different faces or their parts). The last assumption is consistent with the real disposition of cover materials within car cavity. Although such an approximation simplifies real situation it is expected to be available for correct description of the sound field peculiarities.

The ray approach was applied for high frequencies. In the case all the restrictions on materials' disposition are removed. From two possible methods: the ray tracing and the imaginary sources method [6] we applied last one modified in such a manner to account for mutual shading of different parts of a cabin inner surface.

Mathematical formulation of the simulation problem includes an approximation of speech and noise sources. Speech signal source is approximated by point one. It is possible to entertain its directivity at high frequencies.

The noise enters into cabin from outside. Regardless of concrete nature of the noise source (vibrations caused by engine or transmissions, interaction of wheels and road surface, traffic noise) the noise within the cabin is caused by normal oscillation of the cabin surface elements. Due to comparatively high rigidity of shells we can neglect air influence on these oscillations and treat each element as point source [8]. Simplifying the situation we remove the distributed noise source by a finite set of point sources disposed at the respective parts of the cabin surface. These point sources are treated as stochastic ones with specified for each frequency intercorrelation.

After Green's function computing for each pair "point source - MA element" the complex signal spectrum transmitted to an array element is expressed in terms of initial spectrum $S_0(\omega)$ as follows (both for speech and interference):

$$S_p(\omega) = \sum_{m=1}^{M} S_{0m}(\omega) G(\omega, \mathbf{r}_p, \mathbf{r}_m)$$

where $\mathbf{r}_m$ is position vector of the m-th point source, p is the number of an MA element. Then, computed in frequency domain signals and noises at array elements are exposed to the processing in accordance with a concrete algorithm (mentioned in the previous section). Then the obtained spectrums of the signal and noise at MA output are used to reconstruct the output speech and interference in time domain by use of inverse Fourrier transformation.

All the computations necessary for MA optimization were executed by means of the software package "InCarSound" designed for sound fields' simulation within car cavity. This program allows obtaining both theoretical and subjective estimates of the MA effectiveness. Last ones can be obtained by way of audition of the simulated speech and noise mixture before and after processing.

3. NUMERICAL COMPUTATIONS

The typical cabin of an automobile (of sedan kind) was taken into consideration as an example. To simplify model geometry some surfaces were aligned and the cabin was approximated by orthogonal prism of polygonal bases superposed with the side walls of a cabin (see Fig.1). Glass, soft coating of seats, leather coating and coating of the floor were taken as sound absorption materials with their reference characteristics. The noise source was approximated by 13 point sources disposed near engine box.

The effectiveness of the linear MA of length equal 15 cm (parallel or orthogonal to the front wall) consisted of 5 microphones at engine noise background was considered. We estimated the array effectiveness for three versions of spatial-time processing beamforming technique, optimal algorithm and adaptive algorithm of Capon.

Traditional beamforming technique is based on concept of the desired signal coming from the narrow frustum and the interference waves distributed uniformly. Then an array with narrow directivity pattern can filtrate the waves of noise.

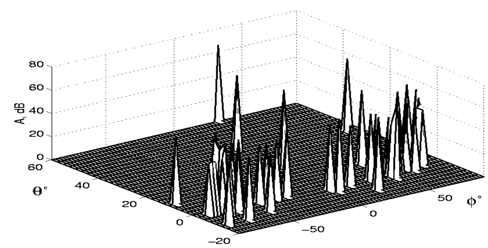

Fig. 2. Distribution of signal amplitudes against angles.

But within a car cavity the signal field structure is more complicated due to numerous reflections. One can see from Fig. 2 that desired signal waves are distributed within comparatively wide frustum closed to that of the waves of interference. Thus one can expect that traditional beamforming technique does not give the effective way to select the signal at noise background.

The optimal algorithm of spatial-time processing (and its adaptive version) is known to be effective on condition of strong noise field correlation at array aperture. Is the correlation sufficiently big? It governs by two factors: cross correlation of point sources and wave propagation law for the specific situation. So it is important to entertain real wave field structure within a car cavity. The greater is the sources' correlation, the greater is the spatial correlation at an array aperture. Different versions of sources' coherence were taken when computing the array effectiveness.

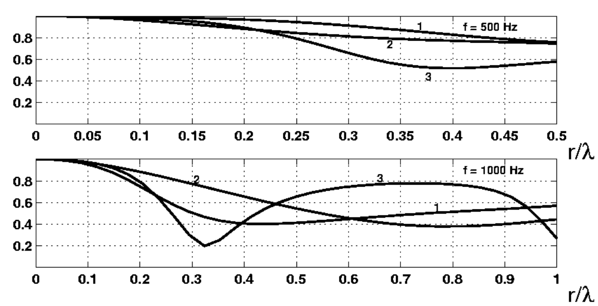

Fig. 3. Correlation at array aperture for different array orientation (1- longitudinal, 2 - vertical, 3 - across the car axis).

For example we suppose noise sources' coherence interval to be frequency dependent: big at low frequencies and small at high frequencies. We can obtain lower-bounded estimate on assumption of noncorrelated point sources.

Computations related to this worst situation are illustrated by Fig. 3 and show that the correlation is sufficiently big (at low frequencies at least). One can expect the optimal algorithm and its adaptive versions to be effective. Presented results point also to the correlation dependence on direction of the array axis. So array orientation (along or across a car axis) is expected to be important.

Computations of the effectiveness parameter (determined by expression (4)) validate above-mentioned conclusions on different algorithm effectiveness (see Fig. 4). All these results relate to the worst case when the point noise sources are assumed to be noncorrelated. The upper figure corresponds to the array oriented across car axis and the lower one corresponds to the case of the parallel to the axis orientation.

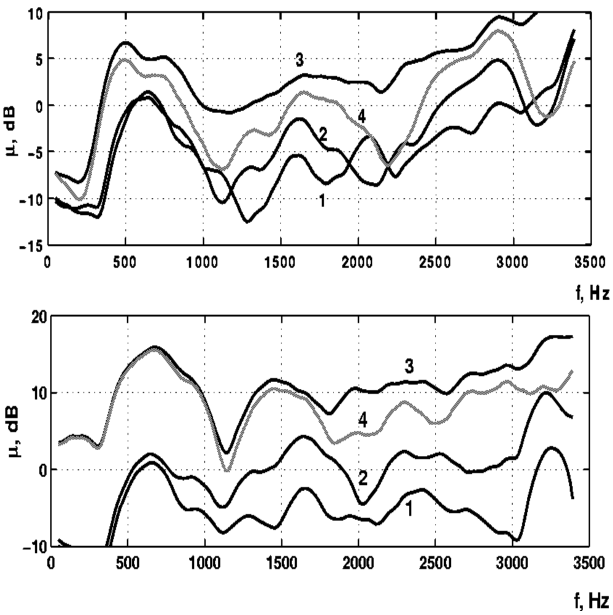

Fig.4. Signal to noise ratio (dB) at output of MA: 1 - single ME , 2 - traditional beamforming, 3 - optimal processing, 4 - Capon's algorithm. Noise sources are presumed to be noncorrelated.

One can see that the traditional beamforming technique does not improve essentially the signal to noise ratio (see curve 2) as compared to single omnidirectional microphone (the curve 1). But the optimal spatial processing with APD as a solution of equation (5) amplifies this ratio significantly (compare curves 3 and 1). Capon's algorithm with APD (7) (curve 4) gives the good results as well. Array orientation along a car longitudinal axis is preferable because of more strong correlation at array aperture for this direction.

Subjective estimates of the algorithms' effectiveness were obtained by audition of simulated mixture of speech and noise signals and by comparing the array input and output signals (i.e., before and after processing). Fig. 4 demonstrates an example of the oscillograms of the pure speech sample, simulated speech at engine noise background and cleaned speech after optimal processing.

5. CONCLUSION

Investigation of small MA effectiveness based on developed mathematical model of sound field within a car cavity show their applicability for speech enhancement in cars. Adaptive versions of the optimal spatial time processing provide effective selection of a voice call at noise background. In the case of engine noise the longitudinal orientation of an array axis is preferable to use.

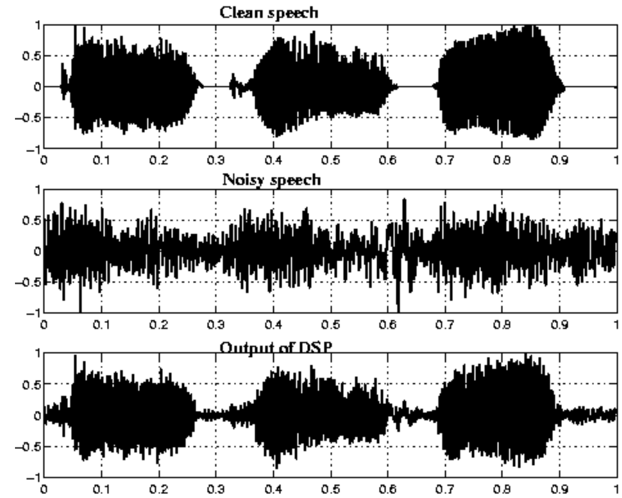

Fig.5. Results of speech enhancement by means of spatial processing. Mixture of test speech signal and interference at the output of single ME (2nd line) is taken in ratio 1/1. Array consists of 5 elements with the optimal APD.

6. REFERENCES

[1] J. L. Flanagan, J. D. Jonston, R. Zahn, G. W. Elko. "Computer steered microphone arrays for sound transduction in large rooms". J. Acoust. Soc. Am., vol.78, pp.1508-1518, 1985.

[2] J. E. Greenberg, P. M. Zurek. "Evaluation of an adaptive beamforming method for hearing aids". J. Acoust. Soc. Am., vol. 91, pp. 1662-1676, 1992.

[3] Zhao Li, M. W. Hoffman. "Evaluation of microphone arrays for enhancing noisy an reverberant speech for coding". IEEE Trans. Speech Audio Processing, vol. 7**,** No. 1, Jan. 1999.

[4] J. Gonzales-Rodriguez, J. L. Sanchez-Bote, J. Ortega-Garsia. "Speech dereverberation and noise reduction with a combined microphone array approach", IEEE Int. Conf. on Acoust., Speech and Signal Proc. "ICASSP 2000". Istanbul, June 2000.

[5] R.A. Monzingo, "Introduction to Adaptive Arrays," John W. & Sons, N.-Y., 1986

[6] M. Dance, B.M. Shield, "The comparison of five computer models for the prediction of sound propagation in industrial rooms," in Collected Papers of 137th meeting of the ASA and 2nd conv. of the EAA. Berlin, 1999.

[7] A. Kovtonyuk, V. Galanenko, A. Kalyuzhny. "Simulation of a Low Frequency Noise Inside a Car Cabin Based On Analytical-Numerical Solution of the Wave Equation", "InterNoise 2000", Nice, August 2000.

[8] C. Floc'h, A. Bardot, J. D. Polack, X. Bohineust. "Vibro-acoustic simulation using geometrical acoustics in the medium range inside car cavity" , "InterNoise 2000", Nice, August 2000.