Engineering Applications of Artificial Intelligence 34 (2014) 37–44

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

An evolutionary algorithm to optimize the microphone array configuration for speech acquisition in vehicles

David Ayllón *, Roberto Gil-Pita, Manuel Utrilla-Manso, Manuel Rosa-Zurera

Department of Signal Theory and Communications, University of Alcala, Alcalá de Henares, Madrid 28508, Spain

ARTICLE INFO

Article history:
Received 16 May 2013
Received in revised form
21 January 2014
Accepted 8 May 2014
Available online 3 June 2014

Keywords:
Microphone array
Speech enhancement
Speech acquisition
Intelligent vehicles

ABSTRACT

Speech acquisition using microphone arrays is included in a variety of trending applications. Multichannel speech enhancement based on spatial filtering aims at improving the quality of the acquired speech. The optimization of the filter coefficients has been the primary focus in beamformer design. However, the array configuration plays an important role in the quality of the speech acquisition system and it should also be optimized. In some applications, the possibilities for microphone placement are very large, and the search of the optimum solution, which involves exploring all possible microphone configurations, is an unfeasible task. This work presents a novel search algorithm based on evolutionary computation to approximate the optimum array configuration. A realistic car noise model based on real measurements is proposed and used in the design. The obtained results support the suitability of the method, notably improving the results obtained by linear arrays with the same number of elements, which are the typical arrays currently assembled in vehicles.

1. Introduction

Nowadays, microphone arrays are useful for a variety of applications where speech acquisition is essential: multi-conference systems (Elko, 1996), monitoring and room control (Busso et al., 2005), hearing aids (Hamacher et al., 2005) or vehicles (Cho and Krishnamurthy, 2003). The common objective of using a microphone array is to enhance the quality of the acquired speech, although the final application of the enhanced speech may differ: speech enhancement in hearing aids, speech recognition in control systems, or voice coding for mobile communications in vehicles. The success of these speech-based applications relies on the quality of the speech acquired by the microphones, which is usually contaminated by different types of noise and interferences. In the last few years, vehicles have become smarter, offering advanced speech-based applications like hands-free telephony, teleconferencing, speech dialog and recognition, allowing the driver to focus on driving instead of handling an electronic device, which increments the driving safety. These advanced features are also demanded in the military environment to enhance vocal communications among military vehicles and the command post. The requirements in military applications are even higher than in civil applications and the environment conditions are more adverse, hence, the robustness of the speech acquisition system is absolutely essential. Some different applications of microphone arrays in the military environment are found in Okandan et al. (2007) and Goldman (2009).

The use of directional microphones pointing towards the desired speech source is an alternative for low-noise speech acquisition, but directional microphones are normally larger and more expensive than omnidirectional microphones, and they do not allow adapting the directivity to the movements of the desired sound source. The use of microphone arrays built of cheap omnidirectional microphones combined with spatial filtering solve these limitations. Beamforming techniques combine the signals collected at the M input channels of the array in such a way that the signals coming from the desired direction are coherently combined and the signals coming from different directions are incoherently combined. There are different strategies to compute the filter coefficients, for instance, the well-known superdirective beamformer (Hansen and Woodyard, 1938), which optimizes the filter coefficients to maximize the array gain. A comprehensive review of beamforming techniques for speech enhancement can be found in Benesty et al. (2008). Nevertheless, the design of the spatial filter is not the only factor affecting the quality of the acquired speech: the position, geometry and number of microphones of the array also have a strong influence in the performance of the system (Feng et al., 2012). Consequently, the selection of the array configuration should be carefully studied in the design. In some applications, the area available to place the microphones is very large, e.g. within a vehicle or a room. This fact implies a large range of possible array configurations, which makes unfeasible an exhaustive search to obtain the best solution.

* Corresponding author. Tel.: +34 918856662
E-mail address: david.ayllon@uah.es (D. Ayllón).

http://dx.doi.org/10.1016/j.engappai.2014.05.006

In this paper, a tailored search algorithm based on evolutionary computation is proposed to approximate the optimum microphone array configuration, in terms of array gain, for speech acquisition in a vehicle. In spite of the fact that different array configurations for vehicles have been proposed during the last few years, for instance, in Ayllón et al. (2012), Grenier (1992), Martin et al. (2001), and Oh et al. (1992), none of them corresponds with an optimized solution but just rough approximations that obtain good results. Evolutionary algorithms have been largely used in engineering to solve optimization and search problems in a wide range of applications, for instance, automatic speech/music discrimination (Ruiz-Reyes et al., 2010), adaptation of non-native speech in a speech recognition system (Selouani and Alotaibi, 2013), antenna array design in Chabuk et al. (2012), or mobile robot localization (Kwok et al., 2006). Additionally, we propose a novel vehicle noise model based on noise recordings under normal driving conditions. The speech inside a vehicle is mainly contaminated by two types of interferences: background noise due to normal driving conditions, which has been usually modeled as a diffuse noise field (Bitzer et al., 1999), and directional signals coming from the loudspeakers. The proposed model is used by the search algorithm to approximate the optimum array.

The remainder of the paper is organized as follows. In Section 2 the proposed car model is described, analyzing the signals recorded in a real car under normal driving conditions. Section 3 describes the signal model and the beamforming technique used in this work. Section 4 describes the proposed search algorithm, and Section 5 contains a description of the experiments carried out in this paper and the obtained results. Finally, Section 6 ends with the conclusions derived from the results.

2. Proposed car model

The background noise due to normal driving conditions inside a car is commonly modeled as a diffuse noise field (Bitzer et al., 1999), which assumes an infinite number of spatially uncorrelated isotropic noise sources. A measure for describing the noise environment is the complex noise field coherence between two signals $x_i(t)$ and $x_j(t)$, with discrete time index t. In the frequency domain, it is defined as (White and Boashash, 1990)

$$\Gamma_{ij}(k) = \frac{\Phi_{x_i x_j}(k)}{\sqrt{\Phi_{x_i}(k)\Phi_{x_j}(k)}}, \tag{1}$$

where $\Phi_{x_i}(k)$ and $\Phi_{x_j}(j)$ represent the auto-power spectral density (PSD) of x_i and x_j, respectively, $\Phi_{x_i x_j}$ represents their cross-PSD, and k represents frequency, $k = 0, \ldots, K-1$. In a diffuse noise field, all microphones receive equal amplitude and random phase noise signals from all directions. In such a case, the coherence between the noise signals acquired by two microphones depends only on the distance between them, according to

$$\Gamma(k) = \mathrm{sinc}\left(\frac{2\pi f d_{nm}}{c}\right), \tag{2}$$

where d_{nm} is the distance between the n-th and the m-th microphone and c is the speed of sound.

In order to validate the assumption of diffuse noise, we have carried out several recordings of the background noise inside a car under normal driving conditions, using an uniform linear array composed of eight omnidirectional microphones with a microphone distance of 35 mm. The analysis of the real noise acquired inside the car reveals that the coherence of that noise differs from the coherence of an ideal diffuse noise field. Fig. 1 shows the spatial coherence of an ideal diffuse noise field with a microphone distance of 35 mm (red line), and the average spatial coherence of the real noise acquired in the car (black line). In the latter case, the coherence has been calculated by averaging the coherence between all possible pairs of consecutive microphones. Although the lower frequencies are highly correlated and the coherence starts to decrease with frequency, for frequencies above 500 Hz the coherence is much higher than the expected for an ideal diffuse noise field. This fact motivates the idea that using a model with a finite number of noise sources is more adequate.

The car model we propose have the following characteristics: (a) the volume of the car is approximated by a cuboid; (b) most of the noise energy inside the car comes from limited areas along its contour, such as windows, wheels and engine; (c) there is a large but finite number of spatially uncorrelated white noise sources distributed along the different areas of the vehicle where noise sources are. Fig. 2 shows the different noise sources considered in this model, represented by black spheres: 15 for the front

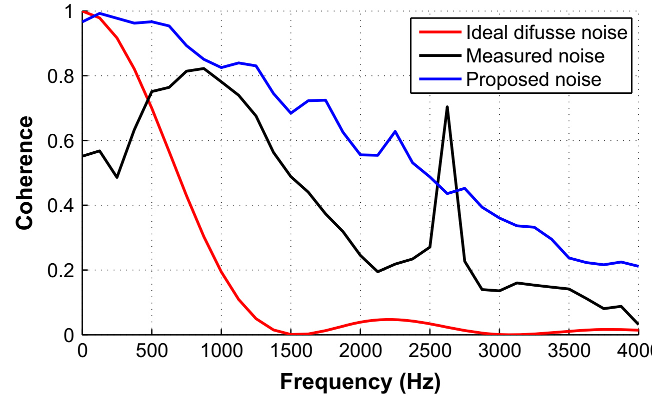

Fig. 1. Spatial coherence of an ideal diffuse noise field (red line), of the real noise acquired in a car (black line), and the proposed noise model (blue line). The coherence has been calculated using two microphones with a distance of 35 mm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

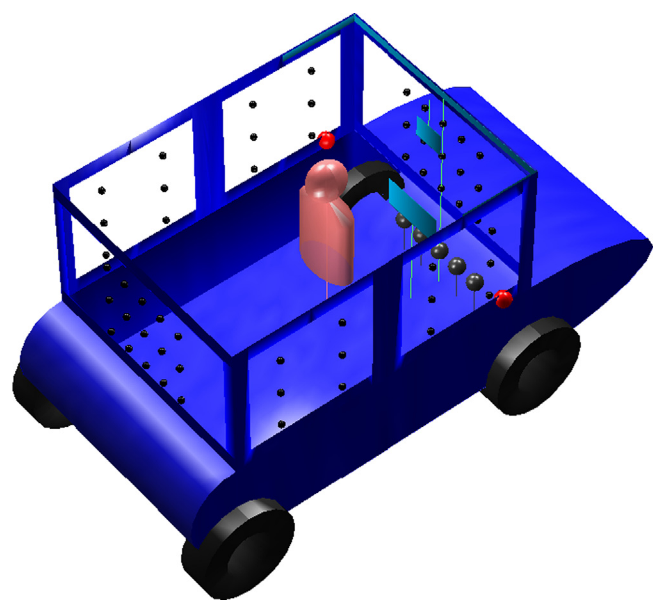

Fig. 2. Car noise model. Black spheres represent the positions of the 63 noise sources, and the two red spheres the front loudspeakers. The target source is located in the head of the dummy driver, and the light blue rectangles represent the areas where the microphone can be placed: the area in the rear-view mirror (A1), the area located in the center of the dashboard (A2), and the upper front edge between the windshield and the roof and the edge between front lateral windows and the roof (A3). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)