Alexander B. BEKIARSKI, Snejana G. PLESHKOVA-BEKIARSKA
FACULTY OF COMMUNICATIONS TECHNIQUES AND TECHNOLOGIES, TU, SOFIA, BULGARIA

Speech Signal Measurement with 2D Microphone Array for Audio Visual Robot Control

Ph.D. Alexander BEKIARSKI

The Author was born in 1944 in Plovdiv, Bulgaria. He received M.S. degree in electronic and automatic engineering from Technical University, Sofia, in 1969. Actually he is Ph.D. associate professor in this University. He published over 180 research papers. His research interests comprise Visual and Audio Systems, Pattern and Speech Recognition, Digital Image and Audio Signal Processors, Neural Networks, Artificial Intelligence, Audio and Visual Robot Systems, etc.

Ph.D. Snejana PLESHKOVA-BEKIARSKA

The Author was born in 1964 in Dobritch, Bulgaria. She received M.S. degree in electronic and automatic engineering from Technical University, Sofia, in 1989. Actually she is Ph.D. assistant professor in this University. She published over 50 research papers. Her present research interests comprise Neural Networks, Signal Processing and Pattern Recognition, Digital Image and Audio Signal Processors, etc.

e-mail: snegpl@tu-sofia.bg

Abstract

Speech signals are one of the essential sources of information in the field of modern intelligent robots, equipped with a microphone array as audio sensors. Applications of microphone arrays are well known. They are used to collect and measure the audio information in audio processing system of a robot. The audio information can be of different nature: music, speech, noise etc. The paper refers only to speech signals, which are used for robot control. There are many structures of the microphone arrays: linear, planar, circular etc., which can be used for collecting and measuring the speech signals with the audio system of an audio visual robot. Most often linear microphone arrays are used mainly because of theirs simplicity. They are also used for robot orientation and movement control in simple room situation, by means of the direction detection of speech arrival. The goal of this paper is presentation of the use 2D microphone array for speech signal measurement, and applying space-time filtering optimized to find speech direction of arrival (DOA). The discovered and calculated speech signal direction of arrival can be combined with the video sensor co-ordinate information to effectively control the mobile robot movements in specified direction.

Keywords: speech signal measurement, audio visual robot control, audio visual robot sensors, microphone arrays, speech processing.

Pomiar sygnału głosowego za pomocą matrycy mikrofonowej dwuwymiarowej przeznaczonej do audio-wizyjnego sterowania robota

Streszczenie

Sygnał mowy jest jednym z głównych źródeł informacji dla współczesnych robotów inteligentnych, wyposażonych w matryce mikrofonowe pracujące jako sensory sygnału audio. Zastosowania takich matryc są dobrze znane. Służą one do zbierania i pomiaru informacji zawartej w sygnałach audio. Informacje audio mogą mieć różną naturę: może to być muzyka, mowa, szum itp. Artykuł dotyczy jedynie sygnałów głosowych, które są używane do sterowania robota. Istnieje wiele struktur matryc mikrofonowych, np. liniowe, planarne, kołowe itd., które mogą być używane do zbierania i pomiarów parametrów sygnału mowy przez system audio robota. Najczęściej z powodu ich prostoty są stosowane matryce liniowe. Wykorzystuje się je również do orientowania robota i sterowania jego ruchem w prostej sytuacji, gdy robot pracuje w pokoju, za pomocą wykrywania kierunku z którego przychodzi sygnał głosowy. Celem artykułu jest przedstawienie zastosowania dwuwymiarowej matrycy mikrofonowej do pomiaru sygnału głosowego oraz zastosowania filtracji czasowo-przestrzennej zoptymalizowanej do znajdowania kierunku z jakiego przychodzi sygnał głosowy (DOA). Wykryty i obliczony kierunek nadchodzenia sygnału głosowego może być połączony z informacjami o współrzędnych z sensora video w celu efektywnego sterowania ruchów robota mobilnego w określonym kierunku.

Słowa kluczowe: pomiar sygnału głosowego, sterowanie audio-wizyjne robota mobilnego, matryca mikrofonowa, przetwarzanie sygnału mowy, sensory robota.

1. Introduction

The audio and visual systems are an inseparable parts of the intelligent moving robots [1]. For a correct and effective control of an audio visual moving robot in different situations in area of observation it is of particular importance to set the condition of all robot systems to work together or as an set. This means, that the perceived and measured information from each robot sensor, for example from audio sensor or microphone array, must be taken into account. One other example is the information from visual robot sensor or video processing system.

2. Sound wave propagation in area of audio visual robot observation

The sound input signals from a pair of microphones or microphone array in the audio robot are a combination from one direct, many reflected sound waves and noise. For a good and precise measurement for source sound localization, it is necessary initially to know the sound waves propagation in area of observation of an audio robot. There are some theoretical studies of acoustics wave propagation in a closed space, usually for musical applications in the concert halls and studio recording rooms [2, 3, 4] and [5]. In [3] a sound propagation model is proposed for sound localization in area of robot observation using two microphones or a linear microphone array. This paper refers to speech signal processing in 2D microphone array, applying the theory of sound waves propagation and defining an appropriate model for sound wave propagation in the area of observation of an audio robot system.

It is assumed that the area of observation is a room, and the medium of sound waves propagation (air) is homogenous. The sound wave equation in this case can be written in a simplified form:

$$\nabla^2 p = \frac{\rho}{\lambda P_0} . \frac{\partial^2 p}{\partial^2 t} = \frac{1}{c^2} . \frac{\partial^2 p}{\partial^2 t}, \tag{1}$$

where:

p – air pressure;
ρ – air density;
λ =1,4 – adiabatic index for air;
c=331,3+0,6T_c m/s – speed of sound in air;
T_c - temperature in degrees Celsius.

Some simplification can be made with the assumption of a monochromatic plane wave. In this case the solution of the general wave equation (1) is:

$$s(x,y,z,t)=Ae^{\mathrm{j}(\Omega t-k_x x-k_y y-k_z z)}=Ae^{\mathrm{j}(\Omega t-\vec{k}\vec{x})}, \tag{2}$$

where:

Ω is the frequency of the monochromatic plane wave;

$$\left|\vec{k}=\frac{\Omega}{c}\right|, \quad \text{or} \quad k_x^2+k_y^2+k_z^2=\frac{\Omega^2}{c^2}. \tag{3}$$

Vector $\vec{k}$ is known as wave number vector in the dimension of rad/m. Also it is possible to define a unit vector $\vec{\xi}$, describing the wave propagation direction:

$$\vec{\xi}=\frac{\vec{k}}{\left|\vec{k}\right|}, \tag{4}$$

Equation (2) can be modified as:

$$s(\vec{x},t)=Ae^{\mathrm{j}\Omega(t-\alpha.\vec{x})}, \tag{5}$$

where:

$$\vec{\alpha}=\frac{\vec{k}}{\Omega}=\frac{\vec{\xi}}{c} \tag{6}$$

The vector $\vec{\alpha}$ is named slowness, because it is reciprocal to velocity. From equation (6) it is possible to find more generalized solutions by considering a waveform:

$$s(u)=\frac{1}{2\pi}\int_{-\infty}^{\infty}S(\Omega)e^{\mathrm{j}\Omega u}\mathrm{d}\Omega, \tag{7}$$

defined as a Fourier transform of $s(t-\vec{\alpha}\vec{x})$:

$$s(t-\vec{\alpha}\,\vec{x})=\frac{1}{2\pi}\int_{-\infty}^{\infty}S(\Omega)e^{\mathrm{j}(\Omega t-\vec{\alpha}\,\vec{x})}\mathrm{d}\Omega \tag{8}$$

The equation (8) represents a superposition of monochromatic plane waves, i.e. a more generalized solution of the acoustic wave equation with nearly any wave shape $s(u)$, if only a well-defined Fourier Transform exists:

$$S(\vec{k},\Omega)=S(\Omega)\delta(\vec{k}-\Omega\vec{\alpha}_0) \tag{9}$$

The equation (9) represents the line $\vec{k}=\Omega\vec{\alpha}_0$ in the space-frequency domain with the amplitude at each point on the line given by $S(\Omega)$.

After this presentation of plane sound wave propagation for speech sound signals in the area of robot observation, it is possible to use these equations for a space-time filtering of the 2D microphone array signals. This filtering serves in the measurement process as a separator of signals of interest (SOI) from the noise and un-useful reflected signals. The correct selection of signals of interest is important for measurement and estimation of direction of arrival (DOA). The information from the measured speech direction of arrival can be added to the co-ordinate measured using information from visual sensor or video camera of the mobile robot. This can help and improve the work of video sensor in finding the direction of the speech sound arrival, combined with visual recognition of a speaker.

3. 2D Microphone Array Speech Signal Filter for Arrival Direction Measurements

The filtering of the received speech signals from 2D microphone array can be represented [6] in space-frequency domain as:

$$Y(\vec{k},\Omega)=H(\vec{k},\Omega)S(\vec{k}-\Omega\vec{\alpha}_0) \tag{10}$$

or in the space-time domain:

$$y(\vec{x},t)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}h(\vec{x}-\vec{\xi},t-\tau)s(\vec{x},t)\mathrm{d}\vec{\xi}\mathrm{d}\tau, \tag{11}$$

where:

$h(\vec{x},t)$ – the impulse filter response;

$H(\vec{k},\Omega)$ – the filter space-response.

It is possible to make this space-time speech signal filtering for 2D microphone array both in time-domain and in frequency domain, or separate.

The time-domain filtering is shown in Fig. 1.

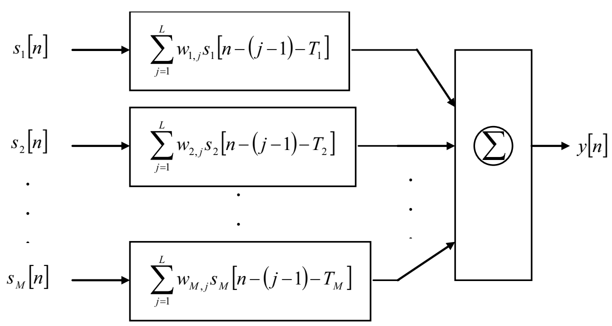

Fig. 1. Block diagram of time-domain filtering
Rys. 1. Schemat blokowy filtru w dziedzinie czasu

The corresponding frequency-domain filtering of 2D microphone array signal is presented in Fig. 2.

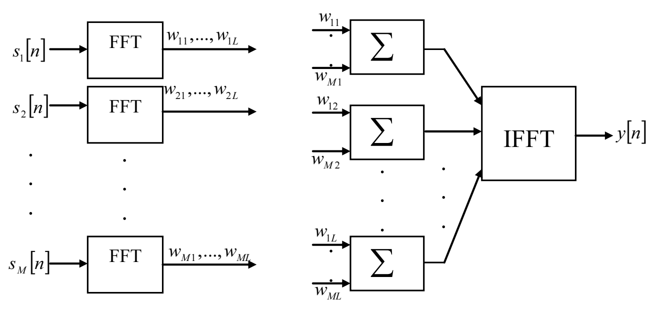

Fig. 2. Block diagram of frequency-domain filtering
Rys. 2. Schemat blokowy filtu w dziedzinie częstotliwości

The corresponding equation of the filter is:

$$y[n]=\sum_{i=1}^{M}\sum_{j=1}^{L}w_{i,j}s_i[n-(j-1)-T_i] \tag{12}$$

where:

$s_1[n],s_2[n],\dots,s_M[n]$ – input signals of 2D microphone array with $M=n_x\times n_y$ – the total number of microphones arranged as a matrix, with dimensions n_x and n_y in horizontal and vertical direction, respectively;

$w_{i,j}$ – weights determined as linear prediction filter coefficients [7] ;
$y[n]$ – the output signal of 2D microphone array.

4. Test results of 2D Microphone Array Speech Signal Filter

The testing of the effectiveness of measured arrival direction of speech signal received from 2D microphone array is taken with National Instruments LabVIEW Sound and Vibration modeling tool [8]. The model is shown in Fig. 3.

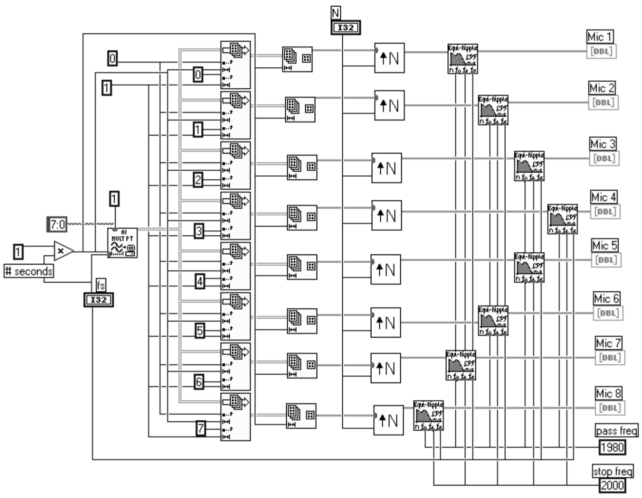

Fig. 3. The National Instruments LabVIEW Sound and Vibration modeling of speech signal filter entered with 2D microphone array

Rys. 3. Model filtru sygnału mowy z matrycą mikrofonową dwuwymiarową na wejściu wykonany przy użyciu karty NI Lab VIEW Sound and Vibration

The number of microphones is eight (0-7) and they are arranged in 4×2 matrix. Some of results from this LabVIEW model are presented in Fig. 4.

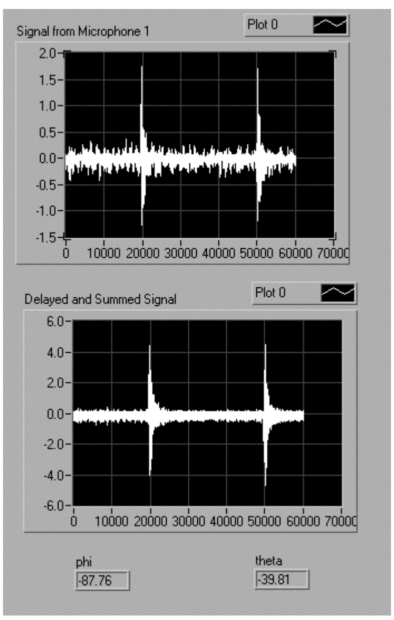

Fig. 4. The input signal from Microphone 1, together with the filter output signal. Measured values for two angles "phi" and "theta" for horizontal and vertical direction of arrival, respectively

Rys. 4. Syganł wejściowy z mikrofonu 1 i sygnał wyjściowy filtru. Sygnały zmierzone dla kątów phi i theta dla kierunku poziomego i pionowego

On the top of this Figure the input signal from Microphone 1 is shown and on the bottom – the output summed signal. Values of measured and calculated two angles "phi = – 87,76°" (in horizontal direction), and "theta = – 39,81°" (in vertical direction), which express 2D direction of arrival (DOA) of speech signal in area of robot observation, are also seen on Fig. 4.

Many tests are made with the proposed measurement tool. All of the results are collected, processed and summarized in the Table 1 as the error between the real and measured angles in percentage for measurements without and with proposed filter.

Tab. 1.

Measurements of 2D Direction of Arrival (DOA)	Error, % (without filter)	Error, % (with filter)
Angle "phi" (horizontal direction)	5%	1%
Angle "theta" (vertical direction)	6%	2%

5. Conclusion

Results achieved from LabVIEW modeling shown that the proposed filtering before measurements of direction of arrival (DOA) of speech signals entered with 2D microphone array are more precise as the measurements without using the proposed filter. In future works it is necessary to perform some additional tests for the precise application of the filter and measurements in a real audio visual robot system, combining the results obtained here with the information measured with visual robot sensors to increase audio visual control, orientation and motion of robot in right direction to the person, speaking in area of observation.

The work was supported by National Ministry of Science and Education of Bulgaria under Contract BY-I-302/2007: "Audio-video information and communication system for active surveillance cooperating with a Mobile Security Robot".

6. References

[1] BEHNKE S.: Humanoid Robots, http://www.hackerz.ir/e-books/ Humanoid Robots.pdf

[2] GILMOR D.E.: Acoustic Source Direction by Hemisphere Sampling, IEEE International Conference on Acoustics, Speech and Signal Processing, vol.2, no.2, pp.28-34, 2001.

[3] VILENAVE D.J. : Systemes de transports intelegents cles en mains, 10-es Recontres INRIA-industrie, pp.138-140, 2005.

[4] MICKIEWICZ W., SAWICKI J. : Spatial audio reproduction by headphones using binaural room impulse response measured individually by the listener. Pomiary Automatyka Kontrola, Vol.53, No 6, pp. 30-33, 2007.

[5] ALEN J.: Image method for efficient simulating small-room acoustics, J.Acoust. Soc. Amer, 65(4), pp. 943-949, 2006.

[6] MILETIEV R., ARNAUDOV R.: Approximation method for valuation of the spectrum response of nonuniformly sampled data, 32nd International Conference Applications of Mathematics in Engineering and Economics, Sozopol, pp. 50-56, 2006

[7] PLESHKOVA S., MOMCHEDJIKOV M.: Fast CELP code book search method. Proceedings of the International Scientific Conference on Energy and Information Systems and Technologies, Sofia, pp. 178-182, 2003.

[8] NATIONAL INSTRUMENTS: NI Sound and Vibration Assistant, 2008, www.ni.com/

Artykuł recenzowany