Expert Systems With Applications 71 (2017) 138–146

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Near field acoustic localization under unfavorable conditions using feedforward neural network for processing time difference of arrival

Marko Kovandžić*, Vlastimir Nikolić, Abdulathim Al-Noori, Ivan Ćirić, Miloš Simonović

University of Niš, Faculty of Mechanical Engineering, Aleksandra Medvedeva 14, 18000 Niš, Serbia

A R T I C L E I N F O

Article history:
Received 23 June 2016
Revised 30 October 2016
Accepted 20 November 2016
Available online 21 November 2016

Keywords:
Processing time difference
Acoustic source localization
Time difference of arrival
Feedforward neural network
Artificial intelligence

A B S T R A C T

Using time difference of arrival (TDOA) is one of the two approaches that utilize time delay for acoustic source localization. Combining the obtained TDOAs together with geometrical relationships within acoustic components results in a system of hyperbolic equations. Solving these hyperbolic equations is not a trivial procedure especially in the case of a large number of microphones. The solution is additionally compounded by uncertainties of different backgrounds. The paper investigates the performance of neural networks in modelling a hyperbolic positioning problem using a feedforward neural network as a representative. For experimental purposes, more than 2000 sound files were recorded by 8 spatially disposed microphones, for as many arbitrarily chosen acoustic source positions. The samples were corrupted by high level correlated noise and reverberation. Using cross-correlation, with previous signal pre-processing, TDOAs were evaluated for every pair of microphones. On the basis of the obtained TDOAs and accurate sound source positions, the neural network was trained to perform sound source localization. The performance was examined using a large number of samples in terms of different acoustic sensors setups, network configurations and training parameters. The experiment provided useful guidelines for the practical implementation of feedforward neural networks in the near-field acoustic localization. The procedure does not require substantial knowledge of signal processing and that is why it is suitable for a broad range of users.

1. Introduction

Acoustic source localization is the determination of sound source position, relative to some reference frame, by using sound signals. It is used in diverse military, industrial, scientific, office, and home applications for speech signal processing (Clifford, Rathborn, & Bull, 1981), intelligent living environments (Principi, Droghini, Squartini, Olivetti, & Piazza, 2016), maintenance and structural monitoring systems (Costiner et al., 2014), sonar (Bokhari & Khan, 2012), surveillance systems (Vozáriková, Pleva, Juhár, & Cižmár, 2011) or to locate sources of artillery fire (Calhoun, Showen, Beldock, Manderville, & Dunham, 2012). The process is performed passively or actively and can take place in liquids, gases and solids. Passive systems use sound signals coming from acoustic sources to locate them. On the other hand, active localization is used for locating targets that are not necessarily acoustic sources. Sound pulses are sent and their echoes, reflected from objects, are used for localization (Bokhari & Khan, 2012).

Many implementations of acoustic localization systems try to imitate flexible and integrated sensory functionality of animals (Nikolić, Kim, & Allen, 2012). People and animals are able to point at the horizontal direction that sound is coming from using slightly different signals that arrive at each ear (Lin, Xiao-Yan, Xu, & Zhen-Yang, 2015). For the vertical direction, features of the sound spectrum, produced by a sound reflector (pinna) (Macpherson & Sabin, 2013), are used as the auditory cue. The localization ability can be established by the process of learning through the repetition of movement. Biologically inspired, audio localization systems can be realized by only one, two or by array of microphones (Argentieri, Danes, & Soueres, 2015; Belloch, Gonzalez, Vidal, & Cobos, 2015; Seewald, Gonzaga, Veronez, Minotto, & Jung, 2014).

Audio localization can be applied on different scales, which mostly depends on the sound power level. Localization of lighting phenomena, volcano explosion (Rowell et al., 2014) or aircrafts (Martín, Genescà, Romeu, & Clot, 2016), for instance, is performed from dozens of kilometers, while in the case of small precision mechanisms or material structure investigation localization is performed on the millimeter scale (Grabowski et al., 2016;

* Corresponding author at: Stanoja Bunuševca 20/82, 18000 Niš, Serbia.
E-mail addresses: marko.kovandzic@gmail.com (M. Kovandžić), vnikolic@masfak.ni.ac.rs (V. Nikolić), abdula.jj@hotmail.com (A. Al-Noori), ciric.ivan@masfak.ni.ac.rs (I. Ćirić), milos.simonovic@masfak.ni.ac.rs (M. Simonović).

http://dx.doi.org/10.1016/j.eswa.2016.11.030

Tan, Zhu, Su, Wang, Wu & Gu, 2016). The error is determined by the geometry of the microphone array, accuracy of the microphone setup, uncertainties in microphones locations, lack of synchronization within microphones, inexact propagation delays, bandwidth of emitted pulses, presence of noise sources, numerical round off errors, anisotropy, obstacles in the propagation path and other terms.

Representative methods of acoustic localization are intensity difference, beam forming and using time delay estimation (TDE). These methods primarily differ in physical variables utilized for the sound source localization. The first method uses the phenomenon of decreasing the sound source energy as it propagates through the medium (Wu, Wang, Dai, & Tong, 2014). Beam-forming uses a collection of signals, from the array of microphones, for computing a correlation matrix, which is thereafter used for the determination of sound source direction (Radcliffe, Naguib, & Humphreys, 2014). The procedure, known as spatial filtering, is based on subspace-theory. The last approach uses the time delay of arrival (TOA), for the case of far-field localization, or the time difference of arrival (TDOA), in the case of near-field localization, collected at different spatial positions. Obtained delays are used, together with the geometrical positions between acoustic components, for estimating the emitter position. Near-field localization is also known as hyperbolic localization because it requires solving hyperbolic equations (Park, Jeon, & Kim, 2014). The complexity of calculations needed for achieving accurate localization increases dramatically with the size of the sensor array (Belloch et al., 2015; Seewald et al., 2014), yet the problem can be mitigated by doing some of the processing on the sensor platform.

The experiment presented in this paper dealt with near field 3D acoustic localization, motivated by the possibility of locating flying objects, such as insects or drones, in the near field of the microphone array on the basis of acoustic signals they emit. It was performed under extremely bad conditions reflected in high disturbances and reverberation. Neural networks were employed for evaluating acoustic source location because of their high speed during exploitation and possibility to model some of the uncertainties in the experimental setup (microphone positions, locations of parabolic reflectors, lack of synchronization within microphones). The performance of neural networks was investigated using a basic, feedforward, neural network, as a representative, in terms of sensor parameters, network configuration and training parameters. The results and the optimal solution of the localization problem are discussed and presented at the end.

2. Time difference of arrival (TDOA)

Sound signals received on two spatially separated audio receivers can be expressed by equations

$$s_i(t) = s_0(t) + n_i(t) \tag{1}$$

$$s_j(t) = \alpha_i s_0(t - \Delta t_{ij}) + n_j(t) \tag{2}$$

where $s_0(t)$ is the signal of emitter, $n_i(t)$ and $n_j(t)$ are the uncorrelated zero-mean Gaussian noise processes, α is the scaled difference in amplitude between the two received signals. After discretization, the previous equations take the form

$$s_i[k] = s_0[k] + n_i[k] \tag{3}$$

$$s_j[k] = \alpha_i s_0[k - l] + n_j[k] \tag{4}$$

where k is the time sample index and l is the correlation lag between the samples. Time difference of arrival, Δt_{ij}, between signals is commonly determined using the cross-correlation function

$$R_{ij}(l) = \frac{1}{K}\sum_{k=0}^{K-1} s_i[k]s_j[k-l] \tag{5}$$

as the argument l that maximizes its value within the range of possible lags

$$\Delta t_{ij} = \frac{1}{F_s}\arg\max\big(R_{ij}[l]\big), \quad -\frac{T}{2} \le l \le \frac{T}{2} \tag{6}$$

where F_s is the sampling frequency and T is the size of the observation window. A good approximation of the cross-correlation function can be obtained using the inverse discrete Fourier transformation

$$R_{ij}(l) \approx \frac{1}{K}\sum_{k=0}^{K-1} R_{ij}(f)e^{\frac{j2\pi fl}{K}} \tag{7}$$

where $R_{ij}(f)$ is the cross-power spectral density (XPSD)

$$R_{ij}(f) = S_i(f)\overline{S_j}(f) \tag{8}$$

Some other TDOA estimation algorithms developed for the purpose of TDOA estimation are phase transform, maximum likelihood estimator, average square difference method, adaptive last mean square filter, etc. All of them differ in accuracy and computational complexity.

TDOA estimation in real circumstances is always negatively affected by disturbances of various backgrounds. The quality of signal is measured by the ratio between the original signal amplitude and the amplitude of noise. It is usually expressed in decibels

$$\mathrm{SNR\ (dB)} = 20\log_{10}\left(\frac{A_{signal}}{A_{noise}}\right) \tag{9}$$

If SNR falls below a certain threshold all methods become unreliable. For the case of the cross-correlation function it is about 13 dB and for the phase transform algorithm about −13.5 dB (Dhull, Arya, & Sahu, 2010).

Since computational efforts have limited effect, SNR ratio is improved using different types of filters. The role of a filter is to suppress the noise while leaving the signal unchanged. The general equation of infinitive impulse response (IIR) filters, in the time domain, has the following form

$$y[n] = \sum_{k=1}^{N} a_k y[n-k] + \sum_{k=0}^{M} b_k x[n-k] \tag{10}$$

Unlike finite impulse response (FIR) filters, the future state of IIR filters depends not only on the finite number of previous inputs, but also on the finite number of previous outputs. The transfer function of IIR filters can be obtained after the z-transformation of Eq. (10)

$$H(z) = \frac{\sum_{k=0}^{M} b_k z^{-k}}{1 - \sum_{k=1}^{N} a_k z^{-k}} \tag{11}$$

A filter design procedure means searching for suitable transfer function coefficients that will provide a filter to meet the specification. For instance, a typical specification of a low-pass filter consists of the following parameters:

- $[0, \omega_p]$ - pass-band
- $[\omega_s, \pi]$ - stop-band
- $[\omega_p, \omega_s]$ - transition band
- $20\log_{10}(1 + \delta_p)$ dB - pass-band ripple (dB)
- $20\log_{10}(\delta_s)$ dB - stop-band ripple (dB)

Graphical representation of these parameters is presented in Fig. 1. Filter design requires considerable knowledge about signal processing and frequency response of both signal and noise. Fortunately, some filters possess the possibility to adapt the transfer function according to an optimization algorithm. Implementing adaptive filters does not require a priori knowledge of signal and noise.