UDC 621.391

Rustem Sinitsyn, assoc. Prof.
Zhanna Bokal
Oleksandra Zaets
Liliya Kolchenko
Anton Tkachuk

NONPARAMETRIC METHOD FOR ESTIMATING THE SPOKEN LANGUAGE SOUND MULTIVARIATE PROBABILITY DENSITY FUNCTION

NAU Radio Electronics Department
E-mail: comandor@i.com.ua

In the paper a new approach for estimating of the spoken language sound multivariate probability density is suggested. It is based on the use of a projection of a random process to the set of random variables, with the probability density defined as a product of two-dimensional densities. The estimates of two-dimensional probability densities are obtained with the help of filtering of the two-dimensional empirical characteristic function. Therefore, we are suggesting a nonparametric estimate of the characteristic function. On the basis of these estimates nonparametric algorithms of sound classification are constructed. Examples for the sound probability density function estimates are suggested.

Запропоновано новий підхід до оцінки багатовимірної щільності ймовірності звуків усної мови, заснований на використанні проекції випадкового процесу на множину випадкових величин із щільністю розподілу ймовірностей, визначених як добуток двовимірних щільностей. Оцінки двовимірних щільностей імовірності отримано за допомогою фільтрації двовимірної емпіричної характеристичної функції. На підставі непараметричної оцінки двовимірної характеристичної функції синтезовано непараметричні алгоритми класифікації звуків. Наведено приклади оцінок щільності ймовірності звукових сигналів, отриманих за пропонованою методикою.

Introduction

It is very important for a spoken language sound recognition to describe a spoken language signal using the most precise and accurate method.

If we consider that a sound signal can be defined as some stochastic process, the fullest description can be obtained in the form of multivariate density function.

Analysis of studies and publications

Using this approach we can solve different problems.

Measurement of a multivariate probability density always was very difficult problem and up to quite a recent time because of complexity of a practical application was not used in a spoken language recognition.

Moreover in mathematical statistics good nonparametric methods of estimating multivariate probability densities have not been developed. The method of multivariate histogram calculation is rather rough and inexact.

Appearance of fast and compact signal processors and kernel methods of estimating of a probability density allows us to find a comprehensible solution of this task now [1].

However, even now this problem is rather difficult.

Problem statement

That is why it is very interesting to suggest new signal models, which simplify this task.

We suggest using a projection of a random process to some simplified model of a multivariate probability density function, which allows us to present the multivariate density function as a product of two-dimensional densities.

For estimating two-dimensional densities, the method of building the kernel estimate of a characteristic function is suggested.

The results of the work are nonparametric estimates of multidimensional probability density and characteristic function.

These estimates can be used for synthesis of nonparametric sound signal detection and recognition algorithms.

The signal model

In the general view, the sound signal can be represented as a random process.

This process is defined in N points of time by the N – dimensional probability density function

$$f(\mathrm{x}) = f_0(x_o) f_1(x_1 \mid x_o) \dots f_N(x_N \mid x_1 \dots x_{N-1}) .$$

This presentation has extreme inconvenience, which reduces its practical application.

We suggest using a projection of a density $f(\mathbf{x})$ to the set of random variables, defined by a probability density

$$f(\mathbf{x}) = f_1(x_1)\prod_{i=2}^{N} f_2(x_i \mid x_1).$$

This projection defines a multivariate probability density as product of two-dimensional probability densities

$$f(\mathbf{x}) = f_1(x_1)\prod_{i=2}^{N} \frac{f_2(x_i, x_1)}{f_1(x_1)}.$$

If we formally suppose that

$$f_2(x_1, x_1) = f_1^2(x_1),$$

we can obtain more simple expression

$$f(\mathbf{x}) = \prod_{i=1}^{N} \frac{f_2(x_i, x_1)}{f_1(x_1)}.$$

The characteristic function of this density is defined by the expression

$$\Theta_N(v_1, v_2, \dots, v_N) =$$

$$= \int_{-\infty}^{\infty} f(x_1)\prod_{i=2}^{N}[\int_{-\infty}^{\infty} f(x_i \mid x_1)e^{iv_i x_i}dx_i]e^{iv_1 x_1}dx_1 =$$

$$= \int_{-\infty}^{\infty} \prod_{i=1}^{N}[\int_{-\infty}^{\infty} \frac{f_2(x_i, x_1)}{f_1(x_1)}e^{iv_i x_i}dx_i]e^{iv_1 x_1}dx_1 =$$

$$= \prod_{i=1}^{N}\Theta_2(v_1, v_i).$$

The kernel estimate of the probability density function based on two-dimensional characteristic function

For the purpose of a signal detection and classification we suggest kernel estimate of the probability density, described in [2]. Also the estimate of two-dimensional probability densities, based on the estimate of the characteristic function [3] can be used (fig.1).

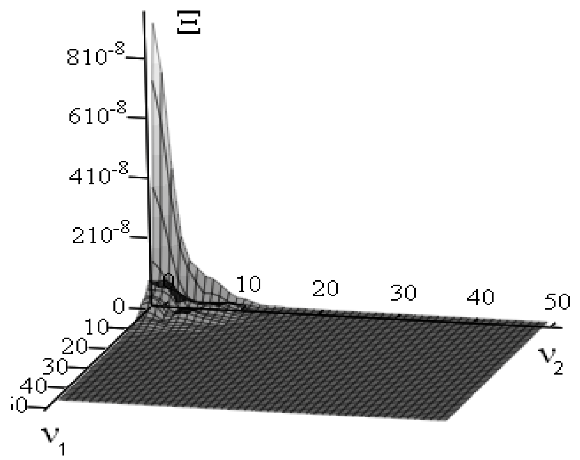

Fig. 1. A kernel estimate of two-dimensional characteristic function for the sound “A”

All units of measurement in fig. 1 are obtained with the help of Fourier transform of the probability density. In this case we can construct two-dimensional empirical probability density function

$$f^*(x_i, x_1) = \frac{1}{M}\sum_{k=1}^{M}\delta(x_1 - x_{1k})\,\delta(x_i - x_{ik}),$$

and two-dimensional empirical characteristic function

$$\Theta^*(v_i, v_1) =$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{M}\sum_{k=1}^{M}\delta(x_1 - x_{1k})\,\delta(x_i - x_{ik})e^{j(v_1x_1 + v_ix_i)}dx_1dx_i =$$

$$= \frac{1}{M}\sum_{k=1}^{M}e^{j(v_1x_{1k} + v_ix_{ik})}, \quad (1)$$

where j is an imaginary unit;

M is the sample size.

The smoothed estimate of the characteristic function can be obtained by multiplication of the empirical characteristic function by the Fourier transform of the kernel

$$\widehat{\Theta}(v_i, v_1) = \Theta_w(v_i, v_1)\Theta^*(v_i, v_1), \quad (2)$$

where $\Theta_w(v_i, v_1)$ is a characteristic function (a kernel of an estimate), corresponding to some probability density $w(\cdot)$:

$$\Theta_w(v_i, v_1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(x_i, x_1)\,e^{j(v_1x_1 + v_ix_i)}dx_1dx_i,$$

$\Theta^*(v_i, v_1)$ is an empirical characteristic function of an estimated statistics.

Use of the characteristic function allows us application of the fast algorithm of convolution calculation with the help of the fast Fourier transform.

The estimate of two-dimensional density is obtained with the help of inverse Fourier transform

$$\hat{f}_2(x_i, x_1) = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\widehat{\Theta}(v_i, v_1)\,e^{-j(v_ix_i + v_1x_1)}dv_idv_1. \quad (3)$$

These estimates for different differences of time are presented in fig. 2, 3, 4.

The estimates are obtained with the help of processing of the acoustic signal for different spoken language sounds.

We can simply prove that this estimate is identically equal to the kernel estimate of a two-dimensional density function

$$\hat{f}_2(x_i, x_1) = \sum_{k=1}^{M} K_k(x_i, x_1),$$

where

$$K_k(x_i, x_1) = \frac{1}{M}w(x_i - x_{ik})\cdot(x_1 - x_{1k})$$

is the kernel of the kernel estimate.

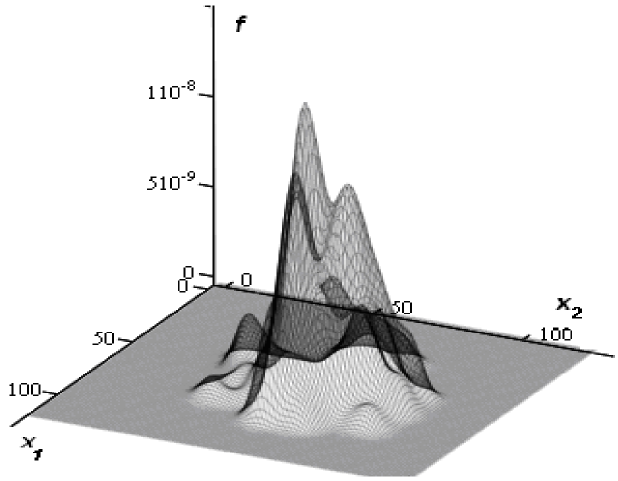

Fig 2. A kernel estimate of the two-dimensional probability density function of the sound "A", horizontal axes represent voltage (ADC digits), the vertical axis represents probability density

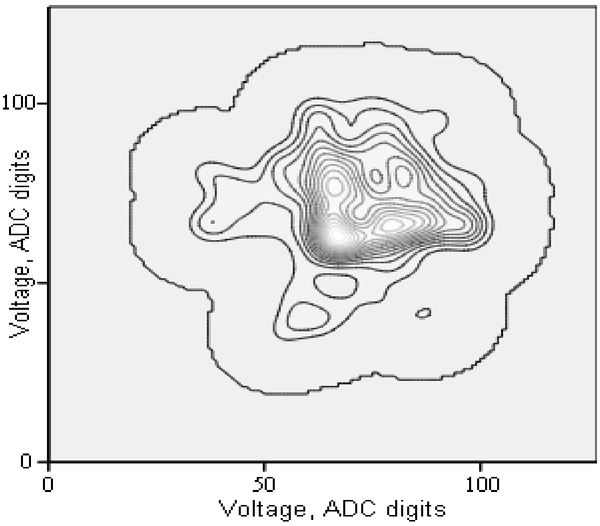

Fig. 3. A kernel estimate of the two-dimensional probability density function of the sound "A" (a view from the top)

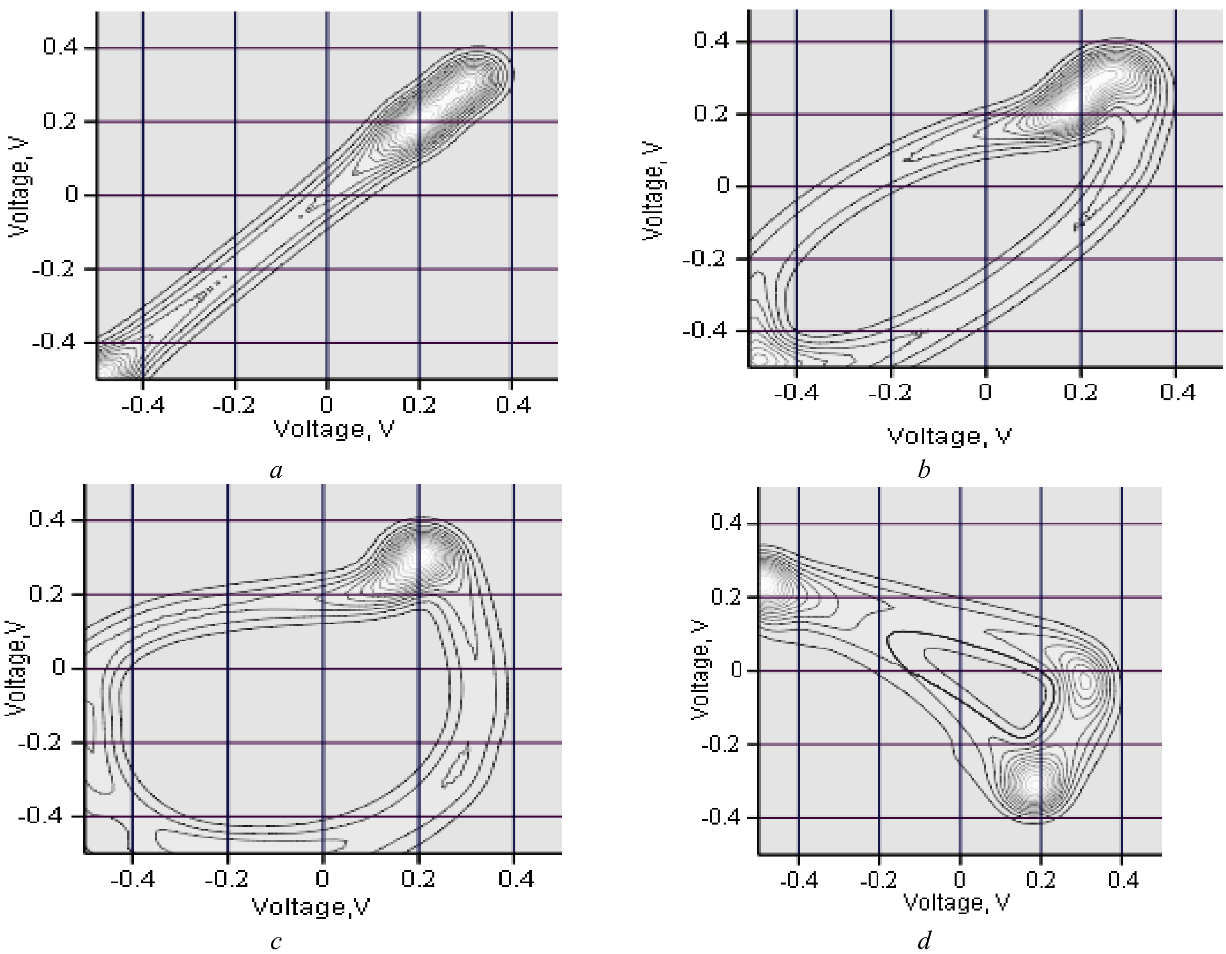

Fig. 4. A kernel estimate of two-dimensional probability density function of the sound signal "U" for two signal values at two moments of time with the interval between them of 0 ms (*a*), 0,417 ms (*b*), 0,833 ms (*c*), 83,3 ms (*d*) (*a* view from the top)

The kernel estimate of the probability density function can be obtained with the help of the expression

$$\hat{f}(x_1, x_2, \ldots, x_N) = $$
$$= \hat{f}_1(x_1)\prod_{i=2}^{N}\frac{\hat{f}_2(x_i, x_1)}{\hat{f}_1(x_1)} = \prod_{i=1}^{N}\frac{\hat{f}_2(x_i, x_1)}{\hat{f}_1(x_1)}$$

or using the estimate of characteristic function we can obtain another variant of the expression with the help of inverse Fourier transform

$$\hat{f}(x_1, \ldots, x_N) = \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\prod_{i=1}^{N}\hat{\Theta}(v_i, v_1)e^{-j(v_i x_i + v_1 x_1)}dv_i dv_1.$$

All expressions can be calculated with the help of the fast Fourier transform (FFT) algorithm. The use of FFT increases the speed of signal processing.

The final expression for multidimensional density is following

$$\hat{f}(\mathrm{x}) = \hat{f}_1(x_1)\prod_{i=2}^{N}\frac{\hat{f}_2(x_i, x_1)}{\hat{f}_1(x_1)}, \quad (4)$$

where $\hat{f}_1(x_1)$ is obtained with a similar method.

Substituting into the estimate of the projection of multivariate density (4) the two-dimensional empirical characteristic function (1), the smoothed estimate of the two-dimensional characteristic function (2) and the Fourier transform of this characteristic function (3) the new expression for the estimate of the probability density function (4) is obtained

$$\hat{f}(\mathbf{x}) = \prod_{i=1}^{N}\frac{1}{M}\sum_{k=1}^{M}\frac{1}{(2\pi)^2}\int_{\infty}^{\infty}\int_{\infty}^{\infty}\Theta_w(v_i, v_1)\times$$
$$\times e^{-jv_i(x_i - x_{ik})}e^{-jv_1(x_1 - x_{1k})}dv_i dv_1. \quad (5)$$

Detection and classification algorithms

Using expression (5) as an estimate of the likelihood function we can design a criterion for multiple decisions using method of the maximum of the likelihood function estimate.

Let us suppose that we have *n* training samples, with the help of which we can build *n* estimates of the probability density function, corresponding to these estimates. We can obtain the decision rule substituting the data from the signal region, which we want to recognize, into our *n* estimates and choosing the maximum among them.

This maximum corresponds to our decision.

Also we can construct ordinary classical signal detection algorithm, using Neumann-Pearson criterion, likelihood ratio and threshold, with which we can compare our decision rule, which can be obtained by substituting of expression (5) instead of likelihood function into the likelihood ratio.

The decision function must be compared with a threshold value.

Some nonparametric invariant algorithms are suggested in [3].

Experiment

The experimental measurements of the multivariate density function were done with help of an ordinary personal computer with the sound card and a microphone. The signal from a microphone is transformed into a digital signal by an analog-to-digital converter (ADC) of the sound card and then processed by the computer. The example of the sound signal is presented in fig. 5.

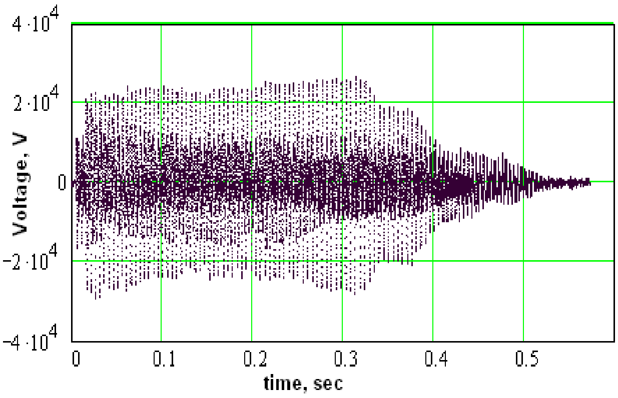

Fig. 5. The sound signal "A"

Conclusions

New nonparametric estimates of the probability density function are suggested and tested in our paper. Our approach allows us to represent estimate of multidimensional density and characteristic functions as a product of two-dimensional probability density function estimates. These estimates can be used for the sound signal detection and recognition. The detection and recognition algorithms are suggested. Real spoken language sound signals are processed and estimates of probability density for these signals are presented.

References

1. *Бабак В.П., Белецкий А.Я., Приставка А.Ф., Приставка Ф.А.* Стохастические сигналы и спектры. – К.: КИТ, 2004. – 290 с.

2. *Sinitsyn R.B., Braun I.M., Yanovsky F.J.,* Nonparametric Detection of Inhomogeneous Areas in Clouds // Conf. Proc. The 11th Conf. on Microwave Technique COMITE 2001, September 18–19, 2001. – University of Pardubice, the Czech Republic. – P. 77–81.

3. *Sinitsyn R.B., Yanovsky F.J.* Kernel Estimates of the Characteristic Function for Radar Signal Detection // European Microwave Week 2005 Conf. Proc., European Radar Conf. 2005. – Paris, October 3–7, 2005. – Horizon House Publications, London. – P. 53–56.

The editors received the article on 3 October 2006.