UDK: 534:681.883(075)
COSATI: 20-01

Analysis of the amplitude distributions of the hydro acoustic river ambient noise using *ARMAsel* algorithm

Miodrag Vračar, MSc (Eng)[1)]

In this work *ARMAsel* algorithm was used for automatic estimation of the *ARMA* parameters of the river hydro acoustics ambient noise to evaluate the nature of the river noise amplitude distributions. Namely, on the base of the estimated *ARMA* parameters digital filter was realized. This filter is used for numeric synthesis of the artificial signals on the bases of various types of numerical distributions, such as: exponential, double exponential, normal, Bernoulli and uniform. The bispectrums were calculated for all of those signals. The bispectrums of artificial signals were compared with bispectrum of the ambient noise signal using two methods and results were systematized in the form of table review. This procedure enabled finding out which bispectrum of artificial signals is closest to the bispectrum of the ambient noise signal. Time duration of all the analyzed signals was in the interval from 100 to 1365 ms.

Key words: **hydroacoustics, ambient noise, *ARMAsel* algorithm, bispectrum.**

Inroduction

THIS analysis is limited to the stationary stochastic processes such as river ambient noise signals of appropriate time duration. Particular analysis of the time series are done using various types of models and algorithms which enables calculations of the power spectral density and covariance function of the various signals [1].

Ambient noise of the river is only one of the possible examples of such time series. Moreover, the ambient noise is considered to be a stationary stochastic process, because the mean value and autocorrelation function change relatively slow in a short period of time [2]. Among the dominant factors those which predominately influence the characteristics of the ambient noise such as wind, rain, hydrology characteristics of the river (flow, speed of water flow, turbulent flow, etc.), biochemical quality, local traffic and another people's activities should be stressed.

Analyzed signals are usually in digit form. Sampling period T_s, might be set to a very small value, compared to T, time period of the signal. This criterion enables obtaining as much information as possible within a limited amount of time or to study both fast and slow phenomena from a single set of data.

Bispectrum analysis is used to estimate the character of amplitude distribution of the ambient noise signal. Namely, amplitude distributions of the river ambient noise signal are compared with five different artificial numerically synthesized signals using bispectrum analysis and special procedure presented in this work.

ARMAsel algorithm is used to automatically estimate the appropriate parameters of the ambient noise signals in order to synthesize artificial signals.

Models of time series

The first half of the last century marks the beginning of work on time series models. It was then that Yule published the first work about auto regressive *(AR)* models. The first description of the real data using moving average *(MA)* models is attributed to Slutsky, but time and frequency domain, were in a unique way independently introduced by Wiener and Khintchine at the beginning of the forties of the last century.

Moreover, it is possible to explain and constitute all the stationary stochastic processes on the bases of one of these three model types to obtain a good spectral description of the data. The main problem is to recognise which type is the most appropriate. Generally speaking, the three mentioned models of time series are:

1. Auto regressive,
2. Moving average and
3. Combination of the two previous models auto regressive moving average *(ARMA)* [3].

Auto regressive model *AR* is realized using discrete linear filter with white noise added on its input. On the output of the filter a random series is generated. Auto regressive model of the p-order is described by the following equation [4]:

$$y(n)+a_1 y(n-1)...+a_p y(n-p)=x(n) \tag{1}$$

For characterization of *AR* processes several parameters are commonly used. A complete description is achieved using poles, reflection coefficients and correlation coefficients.

If output is given as a linear combination of the current and q, the previous values of the signal, the *MA* model is more acceptable and the appropriate equation of the model is:

$$y(n)=x(n)+b_1 x(n-1)+\cdots+b_p x(n-q). \tag{2}$$

Combination the previous models the *ARMA* model is obtained:

$$\begin{aligned} y(n)=&-a_1 y(n-1)-\cdots-a_p y(n-p)+\\ &+x(n)+b_1 x(n-1)+\cdots+b_q x(n-q). \end{aligned} \tag{3}$$

The transfer functions of the filters could be shown using z - transform, while the previous equations are used for solving $y(n)$.

---
1) Military Technical Institute (VTI), Ratka Resanovića, 11132 Belgrade

Models of finite order (p,q) are usually adequate for describing the real processes with infinite orders. This results from the fact that the true parameters decrease for many processes.

The most important question in the implementation of these models would be which of them to use. The answer to this question is extensive and is systematized in the form of the following principals and recommendations:

- if the sharp peaks appear in spectra without expressive zeros the AR model is suitable;
- on the contrary, if the zeros are expressive and without sharp peaks in the spectra, *MA* model is more convenient and
- *ARMA* model is successful when it is necessary to reconcile these two extremes.

It is very important to use the adequate time series model. Furthermore, the application of the wrong model type or order can lead to serious mistakes in spectral estimates. There is no doubt that for all stationary stochastic processes, at least one of the tree model types provides a good spectral description of the data.

ARMAsel algorithm

Selection of the adequate type of the time series model can be done in two ways. The first one is using principal recommendations and the second is the automatic choice using some statistical concept.

In this paper the first approach was not used for the selection of the appropriate model. Namely, the selection of the model type and order was automatic using the so called *ARMAsel* algorithm. *ARMAsel* algorithm enables an optimal choice of the model type and the number of *AR* and *MA* parameters of the signal.

The computational power of the modern computers is on a level that enables using modern algorithms which estimates parametric models of the stochastic time series in an automatic way. From recent times, an automatic selection algorithm was proposed for the selection of the model type *AR*, *MA* or *ARMA*. The proposed algorithm is denoted *ARMAsel*. New criterion in the selection of the model type and order implemented in this algorithm are being used and tested in many fields nowadays.

The automatic *ARMAsel* algorithm has been or is being studied in a number of areas, such as:

- Detection of methacholine from lung sounds;
- Detection of objects in radar clutter;
- Detection of the flow regime in a turbulent flow;
- Spectral representation of irregularly sampled data;
- Spectral estimation with missing data;
- Spectral representation of meteorological data;
- Improved estimate for the accuracy of the estimated mean value, with this accuracy computed from the sum of estimated co-variances;
- Comparison of different spectral models with a single number, the so called model error;
- Extracting the statistically significant peaks from data;
- Giving a brief description of the statistically significant details, which diminishes the number of bits that have to be transmitted, which is important in remote sensing;
- Improving the accuracy of most applications where neutral networks are used in selecting time series models using the *ARMAsel* model that has statistically optimal properties and does not require learning sets where such sets are being used for learning;
- Monitoring the rotating and vibrating equipment;
- In general, monitoring in order to detect statistically significant but un-modeled and unexpected changes in the process behavior in a stochastic environment [4].

In this paper, the *ARMAsel* algorithm was used in a new way. The amplitude distributions of the river ambient noise are analyzed. The noise generated in the river water is a stochastic noise and the consequence of natural and human activities. The main natural factors are turbulent water flow, wind, rain, hail, earthquakes and etc. The most significant contribution to the noise levels is added by the river traffic and other human activities in the vicinity of the place where the ambient noise is observed. The algorithm enabled automatic estimation of the model type and its order. On the bases of estimated parameters, digital filter was realized. To achieve the primary goal it was necessary to synthesize artificial signals used in comparison with the real noise signal. Five artificial signals were used in this analysis. Their amplitudes were distributed in the following way:

1. Single sided exponential distribution;
2. Double sided exponential distribution or Laplace;
3. Normal or Gauss distribution;
4. Bernoulli – Gauss distribution and
5. Uniform distribution.

Further analysis was performed using bispectrums of all those signals.

Bispectrum

It is well known that bispectral analysis is a statistical tool for detecting and identifying nonlinear stochastic signal-generating mechanisms generated in the signal interaction and based on the given data. Bispectral analysis can also be employed to investigate whether the obtained data record is consistent with the hypothesis that the underlying process has Gaussian distribution [5].

Therefore, some authors use bispectrum to test linearity and Gaussianity of the samples. Many previous studies of the ambient noise have been treated as linear processes. However, nowadays it is recognized that several mechanisms contribute to non-linearity of the process. Machinery noise of the vessels can have both nonlinear components and turbulent flow. Thus in the part of the river ambient noise spectrum dominated by shipping is possible to have nonlinear effects. The surface waves also contribute to the appearance of nonlinear components. River waves not such a distinctive phenomenon. Perhaps a more significant phenomenon is the nonlinear coupling between the wind and river water.

The bispectrum is a function of two frequencies. According to some definitions, the bispectrum is two-dimensional Fourier transform of the expected value of the signal at three time points. It would be more precise to say that the bispectrum is defined as Fourier transform of the corresponding cumulant sequence – the third order cumulant. The third order cumulant is defined by:

$$C_{3x}(k,l) = E\{x^{*}(n)x(n+k)x(n+l)\}, \tag{4}$$

where $E\{\ \}$ denotes the expectation operator [6].

Therefore, the bispectrum is defined as two frequencies Fourier transform of the third order cumulant and has a complex value.

$$S_{3x}(f_1,f_2)=\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}C_{3x}(k,l)\cdot e^{-j2\pi f_1 k}\cdot e^{-j2\pi f_2 l} \qquad (5)$$

For process with real values symmetry properties of cumulants lead to symmetry properties of polyspectra. The symmetry properties of the bispectrum can be expressed using the equations given below:

$$\begin{aligned}S_{3x}(f_1,f_2)&=S_{3x}(f_2,f_1)=S_{3x}(f_1,-f_1-f_2)=\\ S_{3x}(-f_1,-f_2,f_2)&=S_{3x}^{*}(-f_1,-f_2)\end{aligned} \qquad (6)$$

Reasons for using cumulants and polyspectra of order $k>2$ can be expressed by the following:

- If $z(n)=x(n)+y(n)$, and $x(n)$ and $y(n)$ are mutually independent processes then the appropriate cumualnts are additive quantities;
- If $x(n)$ is Gaussian then cumulants of higher order $k>2$ equal zero and
- It is possible to recover the higher order cumulants of a non Gaussian signal even in the presence of colored Gaussian noise.

Application of *ARMAsel* algorithm

In this paper bispectra is used for comparison bispectra of ambient noise signal with five bispectra of artificial signals. The comparison is done using two methods. The first method is based on metric of central distances of the absolute value of the bispectra matrix and the second method is conventional, based on quantile-quantile diagrams.

The flow chart of the methods for comparison bispectra is shown on Fig.1.

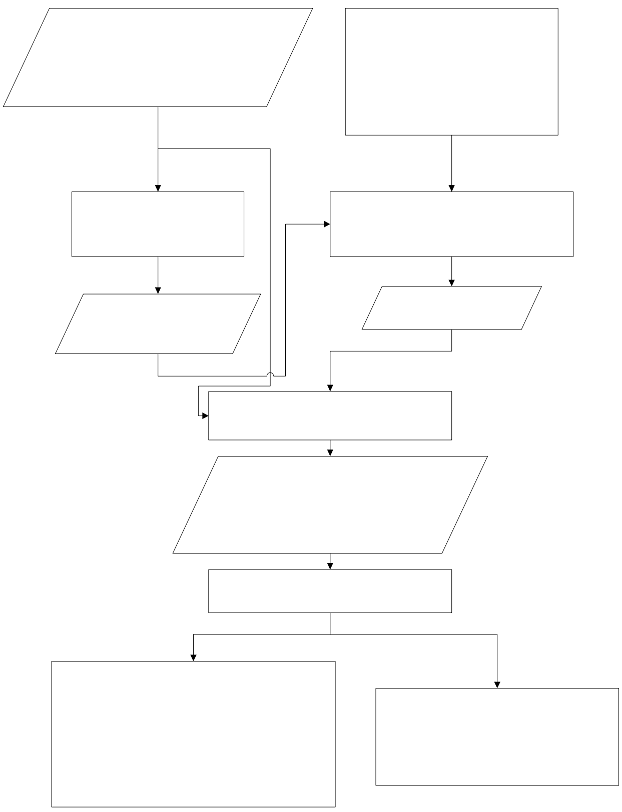

Figure 1. Flow chart of the method used for comparing bispectra of the ambient noise and bispectra of the five artificial signals.

On the bases of these analyses it was possible to determine which artificial signal bispectra is the most similar to the original signal.

Stationarity of the ambient noise

The concept of stationarity is defined on the bases of averaging on the ensemble in specific instants of time. This implies that the existence of multi channel measurements. In practice the derived data have the form of individual time history records of a random phenomenon. When a single time history record is referred to as being stationary this means that the properties computed over short time intervals do not vary significantly during such time intervals.

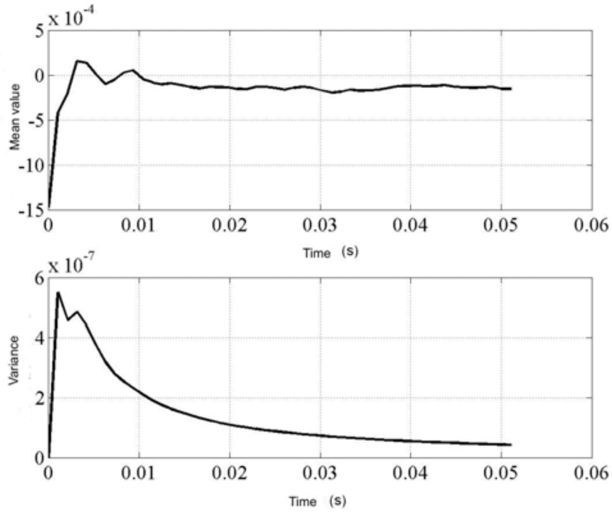

Figure 2. Mean value of the signal amplitude and variance as a function of time.

It is possible to estimate the stationarity of the time process based on the changes of the mean value and autocorrelation function in time. Namely, if the mean value, $\mu_x(t)$ and autocorrelation function $R_x(t,\,t+\tau)$ change in time t, it is said that the process is non-stationary.

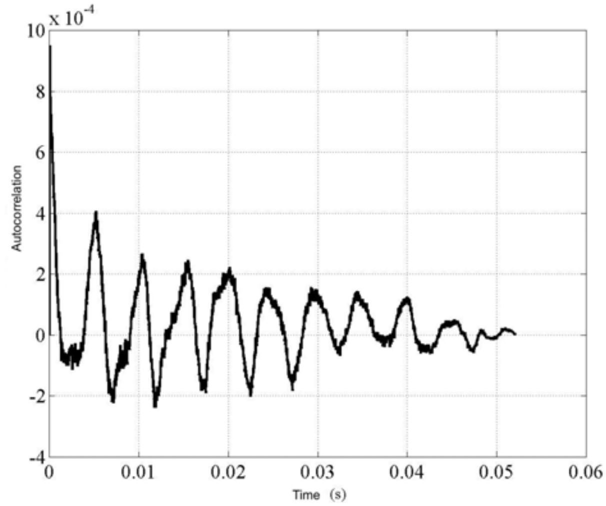

Figure 3. Autocorrelation function of the ambient noise signal

The analysis of the mean value of the signal amplitude, variance and autocorrelation function Figures 1-2, of the typical ambient noise signal lead to the conclusion that the noise signal could be considered stationary after just a short period of time, the order of a few hundred milliseconds.

Results of the application of *ARMAsel* algorithm

ARMA parameters of the ambient noise signal are determined by direct implementation of *ARMAsel* algorithm. On the bases of the calculated *AR* and *MA* parameters, the digital filter, this was used for generating new artificial noise signals was realized. Namely, the five noise signals whose amplitudes were distributed with single-sided exponential, double-sided exponential, normal, Bernoulli and uniform distribution and after that filtered with the outputs used for further analysis. The methodology of the analysis described by flow chart and successive results are shown on Figures 4-6.

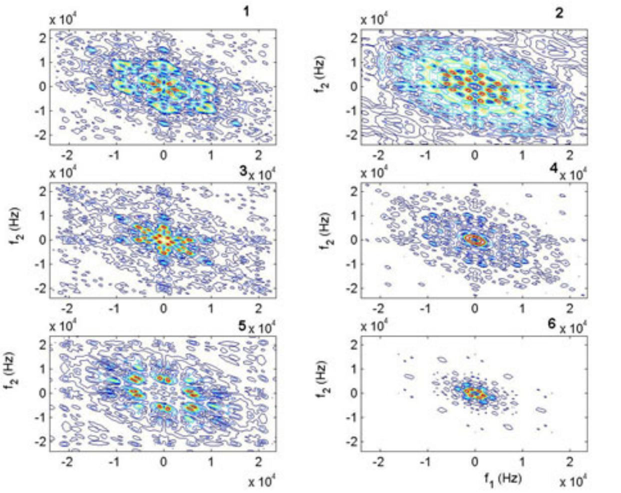

Figure 4. Bispectrums: 1-noise signal, 2–single-sided exponential noise, 3–double-sided exponential noise, 4–normal noise, 5–Bernoulli distributed noise and 6–uniform noise

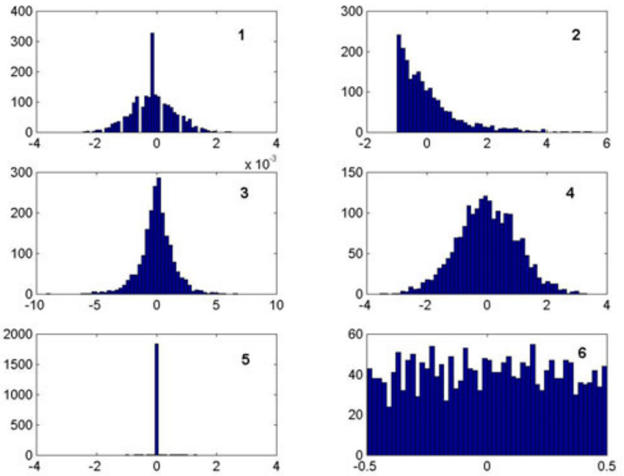

Figure 5. Histograms of the signal amplitudes: 1 – noise signal, 2 – single-sided exponential noise, 3 – double-sided exponential noise, 4– normal distributed noise, 5 – Bernoulli distributed noise and 6 – uniform noise.

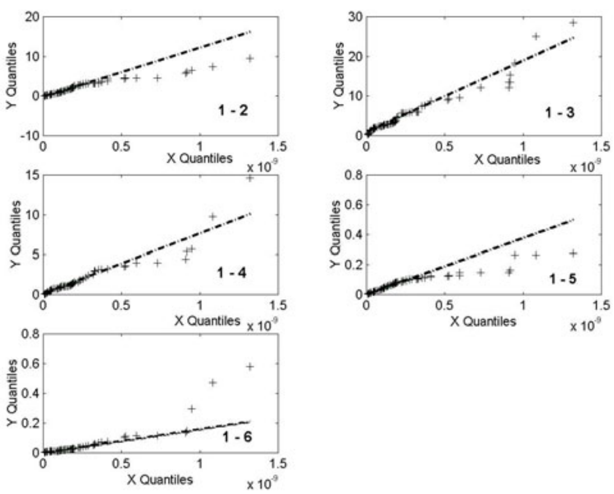

Figure 6. Quantile-Quantile diagrams: 1 – 2 (ambient noise – single-sided exponential noise), 1–3 (ambient noise – double-sided exponential noise), 1 – 4 (ambient noise – normal distributed noise), 1 – 6 (ambient noise – Bernoulli distributed noise) and 1 – 6 (ambient noise – uniform distributed noise)

Comparing the bispectra is done using two methods, one of them is shown on the Fig.6 and is based on the so called *Quantile-Quantile* diagrams (*Q-Q method*) and the other is based on the metric of central distances *(MCD)* of the absolute values of the normalized bispectra matrix, as it was pointed out in the introduction. Namely, the center of the bispectra matrix is taken as the zero point of the coordinate system from which all the distances to other matrix points were calculated and in the end from their sum the measure of the whole bispectral matrix was obtained. The resulting total distance as a measure was suitable for comparing the bispectral matrices realized in the process described in the flow chart (Fig.1).

The analysis is done with the signals of the ambient noise of variable duration. Namely, the noise signal was analyzed in 512, 1024, 2048, 4096, 8192, 16384, 32769 и 65536 points, i.e. in time period from 100 to 1365 ms, starting from the same point. The results of the analysis are systematized in Table 1.

Table 1. The results of analysis: Q-Q diagrams and metric of central distances (MCD) absolute values of normalized bispectra

Number of points		ambient noise signal	single sided exponential distribution	double sided exponential distribution	normal distribution	Bernoulli distribution	uniform distribution
512	*MCD*	0.0046	0.0031	0.0036	0.0046	0.0048	0.0038
	Q-Q					yes	
1024	*MCD*	0.0045	0.0026	0.0051	0.0038	0.0035	0.0063
	Q-Q			yes			
2048	*MCD*	0.0042	0.0026	0.0044	0.0047	0.0035	0.0064
	Q-Q			yes			
4096	*MCD*	0.0047	0.0035	0.0038	0.0039	0.0035	0.0041
	Q-Q						yes
8192	*MCD*	0.0051	0.0023	0.0056	0.0041	0.0041	0.0037
	Q-Q					yes	
16384	*MCD*	0.0048	0.0023	0.0039	0.0053	0.0033	0.0044
	Q-Q						yes
32768	*MCD*	0.0039	0.0018	0.0035	0.0036	0.0040	0.0043
	Q-Q					yes	
65536	*MCD*	0.0017	0.0005	0.0011	0.0012	0.0016	0.0012
	Q-Q					no	

Conclusion

The obtained results show that, in most cases, the model of Bernoulli distribution of amplitudes describes the amplitudes of the river ambient noise in the best possible way. This result is logical because the river is shallow, the water surface relatively small with only one dimension – length, really emphasized. Furthermore, the water ambient is a low frequency filter for acoustics waves. The most probable outcome while measuring ambient noise would be to register direct wave, because the reflected waves are significantly dulled. This analysis also produced some different results, namely the double sided exponential and uniform distribution. This indicates that the problem of amplitude distributions analysis of the river ambient noise is a complex problem with emphasized time and space dimensions.

References

[1] PIET,M.T.,BROERSEN: *Automatic Spectral Analysis With Time Series Models*, IEEE Transacitions on instrumentation and measurements, April 2002, Vol.51, No.2, pp.211-216.

[2] JULIUS,S.,BENDAT and ALLAN,G., PIERSOL: *RANDOM DATA: analysis and measurment procedures,* Wiley-Interscience, 1971.

[3] DUŠAN,B., RAJIĆ,D.: *Uvod u statističku teoriju telekomunikacija*, Akademska misao, Beograd, 2003.

[4] BRORENSEN,P.M.T.: *Facts and Fiction in Spectral Analysis*, IEEE Transacitions on instrumentation and measurements, August 2000, Vol.49, No.4, pp.766-772.

[5] PATRICK,L., BROCKETT,MELVIN HINICH and GARY R.WILSON: *Nonlinear and non-Gaussian ocean noise*, J.Acoust.Soc.Am. 82 (4), october 1987, pp.1386-1394.

[6] Statistics Toolbox for Use with MATLAB®, User's Guide *Version 2.1*.

Received: 01.09.2004.

Analiza amplitudskih raspodela hidroakustičnog rečnog ambijentalnog šuma pomoću *ARMAsel* algoritma

U ovom radu korišćen je *ARMAsel* algoritam za automatsku ocenu *ARMA* parametara rečnog hidroakustičnog šuma u cilju dobijanja ocene raspodelama amplitude šuma. Naime, na bazi dobijenih *ARMA* parametara realizuje se digitalni filtar koji se koristi za numeričku sintezu veštačkih signala na bazi numeričkih raspodela, kao što su: eksponencijalna, dvostruka eksponencijalna, normalna, Bernulijeva i uniformna. Izračunavaju se bispektri svih tako dobijenih signala. Bisplektri veštačkih signala se porede sa bispektrom signala ambijentalnog šuma pomoću dve metode,a dobijeni rezultati su sistematizovani u formi tabele. ova procedura je omogućila da se odredi koji bispektar veštačkih signala je najpribližniji bispektru signala ambijentalnog šuma. Vremensko trajanje svih analiziranih signala je bilo u intervalu od 100 do 1365 ms.

Ključne reči: **hidroakustika, ambijentalni šum, *ARMAsel* algoritam,bispektar**

Analyse des répartitions d'amplitude du bruit hydroacoustique fluvial ambiant à l'aide de l'algorithme ARMAsel

Pour effectuer l'évaluation par répartitions d'amplitude du bruit dans ce travail est utilisé l'algorithme ARMAsel en vue d'une évaluation automatique des paramètres ARMA du bruit hydroacoustique fluvial. A partir des paramètres ARMA obtenus on réalise un filtre digital qui est employé pour la synthèse numérique des signaux artificiels. Cette synthèse est basée sur les répartitions numériques, telles que: exponentielle, double exponentielle, normale, celle de Bernuli et uniforme. On a calculé lles bispectres de tous les signaux reçus de cette manière. Les bispectres des signaux artificiels sont comparés avec le bispectre du bruit ambiant à l'aide de deux méthodes. Les résultats obtenus sont systématisés en forme de tableaux. Ce procédé a permis de déterminer quel bispectre des signaux artificiels est le plus proche du bispectre du signal du bruit ambiant. La durée temporelle de tous les signaux analysés se situe dans l'intervalle de 100 à 1365 ms.

Mots clés: **hydroacoustique, bruit ambiant, algotithme ARMAsel, bispectre**

Анализ амплитудных распределений гидроакустического речного шума окружающей среды при помощи *ARMAsel* алгоритма

В этой работе использован *ARMAsel* алгоритм для автоматической оценки *ARMA* параметров речного гидроакустического шума с целью получения оценки о распределениях амплитуды шума. Иначе говоря, на основании полученых *ARMA* параметров формируется цифровой фильтр, который пользуется для численного синтеза искусственных сигналов на основании численных распределений, каковы: экспоненциальное, двухкратное экспоненциальное, нормальное, Бернулли распределение и форменное. Вычисляются биспектры всех таким образом полученых сигналов. Биспектры искусственных сигналов сравниваются с биспектром сигнала шума окружающей среды при помощи двух методов, а полученые результаты систематизированы в виде таблицы. Эта процедура обеспечила определить, который из биспектров искусственных сигналов ближайший биспектру сигнала шума окружающей среды. Продолжительность всех анализированных сигналов находилась в интервале с 100 по 1365 мс.

Ключевые слова: гидроакустика, шум окружающей среды, ***ARMAsel*** алгоритм, биспектр