Algorithms and results of streaming whistler recognition

Vladimir Mochalov[*], *Anastasia* Mochalova

Institute of Cosmophysical Research and Radio Wave Propagation of the Far Eastern Branch of Russian Academy of Science, 684034, Kamchatskiy kray, Paratunka, Mirnaya str. 7, Russia

Abstract. Algorithms for streaming whistler recognition are offered. Different stages of algorithms are considered. The developed algorithms are used on a mini-computer software and hardware complexes for monitoring very low-frequency electromagnetic radiation at the Karymshina station in Kamchatka, on Oybenkel Geophysical Observatory of the Yu.G. Shafer Institute of Cosmophysical Research and Aeronomy (the Sakha Republic (Yakutia)), as well as recognition of whistlers is performed on the basis of open access data of abelian.org VLF station network in Todmorden United Kingdom (53,703N, 2.072W), Bielefeld Germany (52.146N, 8.458E), Cumiana Italy (44.96N, 7.42E), Warsaw Poland (52.16313N, 21.03094E), Australia Heathcote (36.804163S 144.67559E). Our whistler recognition results allowed us to establish on the days of strong whistler activity, there is an average positive correlation between the number of whistlers registered at Karymshino station in minutes of a day and the number of lightning strokes registered by WWLLN global network in minutes of a day in the coordinate rectangle LAT 25S-45S, LON 140E-160E (Australia).

1 Introduction

In paper [1] the new software-hardware complex «Sensor signal analysis network» (SSAN) for construction and functioning of sensor networks for distributed analysis of time synchronized signals is described. On the basis of the SSAN complex in November 2017 the synchronous registration of atmospherics and whistlers at the radio physical observation stations Karymshino (Kamchatka Krai, Russia) and Oybenkel (Yakutsk, the Republic of Sakha (Yakutia), Russia) began in the operational mode [2]. At each point, a vertical asymmetrical electric dipole as a capacitive sensor is used to register electromagnetic radiation. Signals after amplification are fed to the input of a sound card with a sampling frequency of 192 kHz per channel, which made it possible to see in the spectrograms frequencies of up to 96 kHz and that in turn allowed the registration of big ammount of high-frequency whistlers at the Karymshino station (examples in Figure 1).

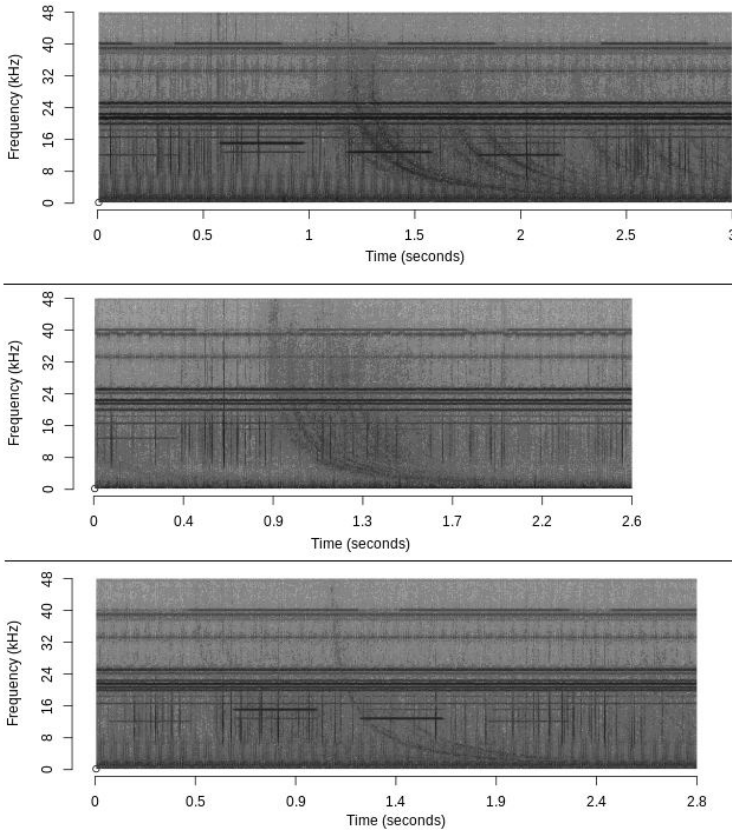[*] Corresponding author: vmochalov@ikir.ru

Fig. 1. Examples of high-frequency whistlers detected and registered with the SSAN complex.

In the SSAN complex, the time reference is made using the GPS module (NMEA protocol and PPS signal). Nodes of SSAN network allow the user to run custom signal analysis programs and to create archives of recognized events, source signals and their spectrograms [1].

With the support of the VarSITI grant, a database of atmospherics and whistlers registered in the Far East of Russia was created [3]. The database was obtained by recognizing events in both archived WAV files and by streaming event recognition using the SSAN complex in near real time mode.

The developed algorithms are also used by the authors for streaming whistler recognition based on the open access data of the VLF station network (abelian.org) in Todmorden United Kingdom (53,703N, 2.072W), Bielefeld Germany (52.146N, 8.458E), Cumiana Italy (44.96N, 7.42E), Warsaw Poland (52.16313N, 21.03094E), Australia Heathcote (36.804163S 144.67559E).

The work is devoted to the description of the developed algorithms and some obtained results obtained on the basis of the application of whistlers recognition algorithms.

2 Existing algorithms for whistlers recognition

At the present time, various algorithms for whistlers recognition [4-7, etc.] have been developed, which distinguish various properties of the presence of whistlers in spectrogram. In paper [4] the authors propose the algorithm developed to determine

plasmaspheric electron density measurements from whistler traces, based on a Virtual (Whistler) Trace Transformation, using a 2-D fast Fourier transform transformation.

The abelian.org website allows to download the open source program vtevent [5], which uses a combination of principal component analysis and Hough transform to detect VLF signals having reasonably narrow bandwidth along with rising or falling frequency. Falling frequency signals are identified as whistlers if they conform to the expected shape of a whistler curve as defined by the low frequency Eckersley approximation. A dispersion range of 12.2 to 114 is examined by the Hough transform.

In the algorithm of [6] the following 4 main stages of whistler recognition are distinguished:
1. median filtering (the authors consider the problem of whistlers recognizing from the point of view of recognizing a graphic image and suppressing extended impulse noise in images suggest using median filtering);
2. selection of significant samples for each column of the filtered spectrogram (the squares of the column elements are considered as the Schuster periodogram of the corresponding fragment of the VLF signal, a check is performed at the significance level α of the hypothesis that this fragment of the signal is white noise);
3. transformation of coordinates for the purpose of straightening the image of the whistler (for the whistler that came at the time τ (the form of the whistler is described by equation $t - \tau = D / \sqrt{f}$), where the coefficient $D > 0$ is called the variance), a transformation is introduced $s = f^{-1/2}$ – then in the coordinate plane (t, s) the whistler image is straightened);
4. recognition of the "straightened" whistler (the whistler search problem is reduced to the problem of finding a set of points near the inclined line (or lines) in the *(t, s)* plane).

In paper [7], an algorithm for automatic recognition of whistlers is proposed, based on the using of the connected component labeling (CCL) method. During the search for whistlers on a spectrogram, the authors remove noise by a number of image processing methods (median filtering, adaptive thresholding and opening methods). After removing the noise, the authors apply the CCL method, which is used to detect the waves pattern in whistler waves spectrogram image. The authors use the CCL method to scan all pixels of the spectrogram image from the upper left to the lower right. Connected authors call the pixels that share the same amount of intensity values and is located closed to each other. After all the connected groups of pixels are highlighted in the image then the groups of pixel can be tagged or marked to distinct one group with others. In the CCL method used, the authors select 2 stages: at the first stage, each pixel is assigned a label in accordance with the four rules (where A(x,y) is the value of pixel intensity of the location (x,y) from image A) [7]:
1. If both the upper neighbor *A(i-1,j)* and left neighbor *A(i,j-1)* of the object pixel *A(i, j)* have the same label X, then assign label X to *A(i,j)*.
2. If either the upper neighbor *A(i-1,j)* or the left neighbor *A(i,j-1)* of the object pixel *A(i,j)* has the label X, then assign label X to *A(i, j)*.
3. If the upper neighbor *A(i-1,j)* has label X and left neighbor *A(i,j-1)* of the object pixel *P(i,j)* has a different label Y (i.e., $X \mathrel{!=} Y$) then assign label X to *A(i,j)*. Enter X and Y in an equivalence table, say it table T.
4. If the upper neighbor *A(i-1,j)* and left neighbor *A(i,j-1)* of the object pixel *A(i,j)* are both not labeled pixels, then assign new label Z to the pixel *A(i,j)*. Enter label Z in the equivalence table T.

Table T contains a set of equivalent labels. At the second stage the equivalent labels are merged to create unique labels for each connected component in the image [7].

3 Functional scheme and new algorithms for whistlers recognition

In order to use the various classification characteristics of the presence in the spectrogram of whistlers and different recognition algorithms, in Figure 2 an extensible functional scheme of the system for whistlers recognition is proposed.
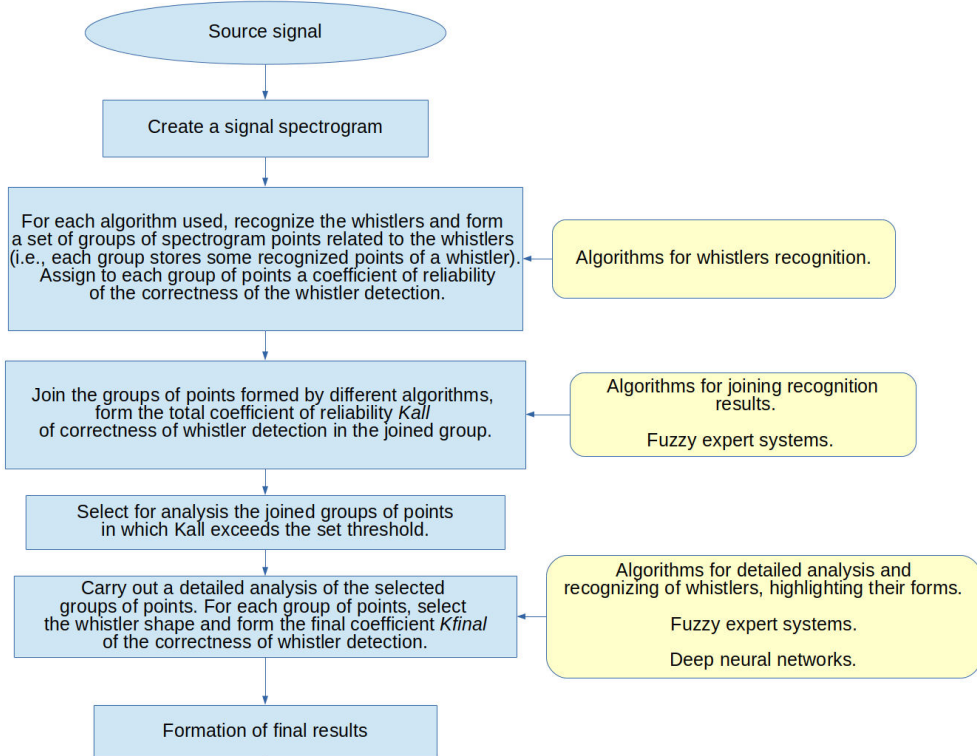

Fig. 2. An extensible functional scheme of the system for whistlers recognition.

Consider a new quick algorithm for whistlers recognition "pull the whistler by the tail". Configurable parameters of the algorithm:
• threshold *Fp*,
• coefficients *Kr* и *Kb* and the array of numbers *I*,
• coefficients *Kr2* и *Kb2* and the array of numbers *I2*,
• coefficient *Ka2*,
• коэффициент *Kg*,
• the length of the rectangle sides *left* and *up*,
• threshold *Pdifx*,
• thresholds *kmin* and *kmax* of the slopes of groups,
• maximal length along the time axis *maxDifX* between the groups to be joined, maximal absolute difference between slope coefficients of groups *kAbsDif*, interval *[difY1; difY2]* of allowable distances along the frequency axis between the left and the right one of the joined groups,
• minimal number of points in the joined group *pJoinedMin*, thresholds *kminJoined* and *kmaxJoined* of slope coefficients of the joined group, a set of required distribution parameters *distJoined* of points in the joined group (for instance, dispersion and

mathematical expectation of the group points deviation from the approximating straight line, distributions and parameters the left and the right halves of group points etc.),

• tables *TLTRD* and *TRTLD* of matching the group size to allowable thresholds of the percentage of the number of group points falling on the covered distance among the total number of group points.

Step 1. Create two-dimensional array *Sl* and write into it the logarithmically normalized values of the signal spectrogram (Figure 3).
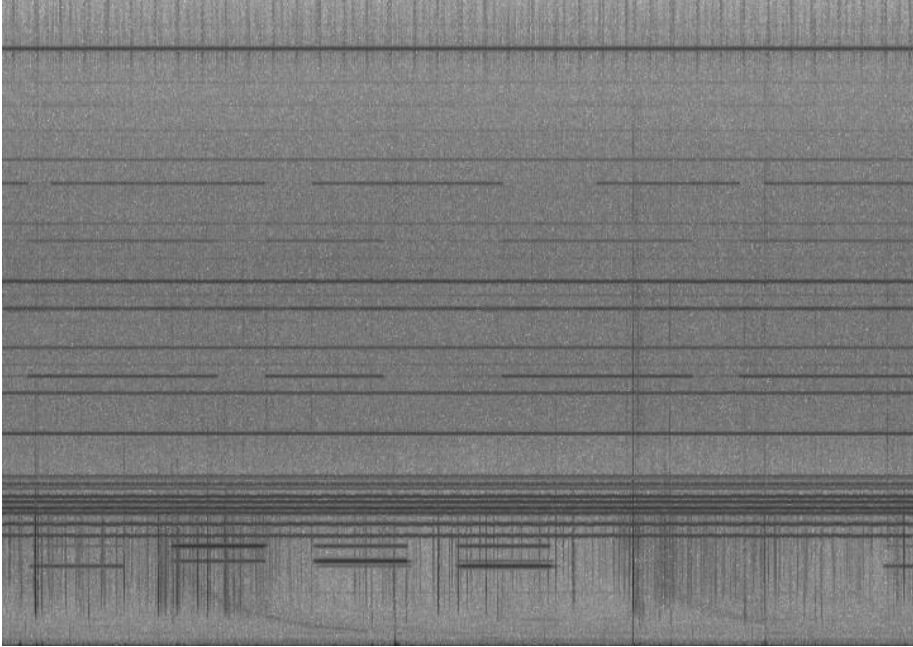
Fig. 3. An example of the step 1 result.

Step 2. Set to zero all values of the spectrogram array *Sl* having frequencies above the given threshold *Fp* (Figure 4(step 2)).

Step 3. Calculate average values of array *Sl* along the frequencies axis and write them into array *Mavg1*. For each element of array $Sl(i,j)>Kr*Mavg1(j)$, for allowable values of its indices ($j+k$), check the condition $Sl(i,j)>Kb*Sl(i,j+k)$ for each $k{\in}I$, and if the condition does not hold for at least one $k{\in}I$, then set $Sl(i,j)=Mavg1(j)$ (Figure 4(step 3)).

Step 4. On the basis of *Sl*, create a spectrogram *Slfreq* with logarithmic frequency axis and putting into *Slfreq* only the frequencies above 200 Hz. Calculate average values of the array *Slfreq* along the frequency axis and write them into array *Mavg2*. For each element of array $Slfreq(i,j)>Kr2*Mavg2(j)$, for allowable values of its indices ($j+k2$), check the condition $|1-\dfrac{Slfreq(i,j)}{Slfreq(i,j+k2)}|<\text{Kb}2*Slfreq(i,j)$ for each $k2{\in}I2,$ and if the condition does not hold for at least one $k2{\in}I2,$ then set $Slfreq(i,j)=Mavg2(j)$.

For each element of array $Slfreq(i,j)<Ka2*Mavg2(j)$, set $Slfreq(i,j)=0$.

Step 5. Create groups of points *G* of the array *Slfreq* using the following rules:

A. Add different points A$(i1,j1)$ and B$(i2,j2)$ (where $|i1-i2|<left$ and $|j1-j2|<up$) into one group if $Slfreq(i1,j1)>Kg*Mavg2(j1)$ and $Slfreq(i2,j2)>Kg*Mavg2(j2)$.
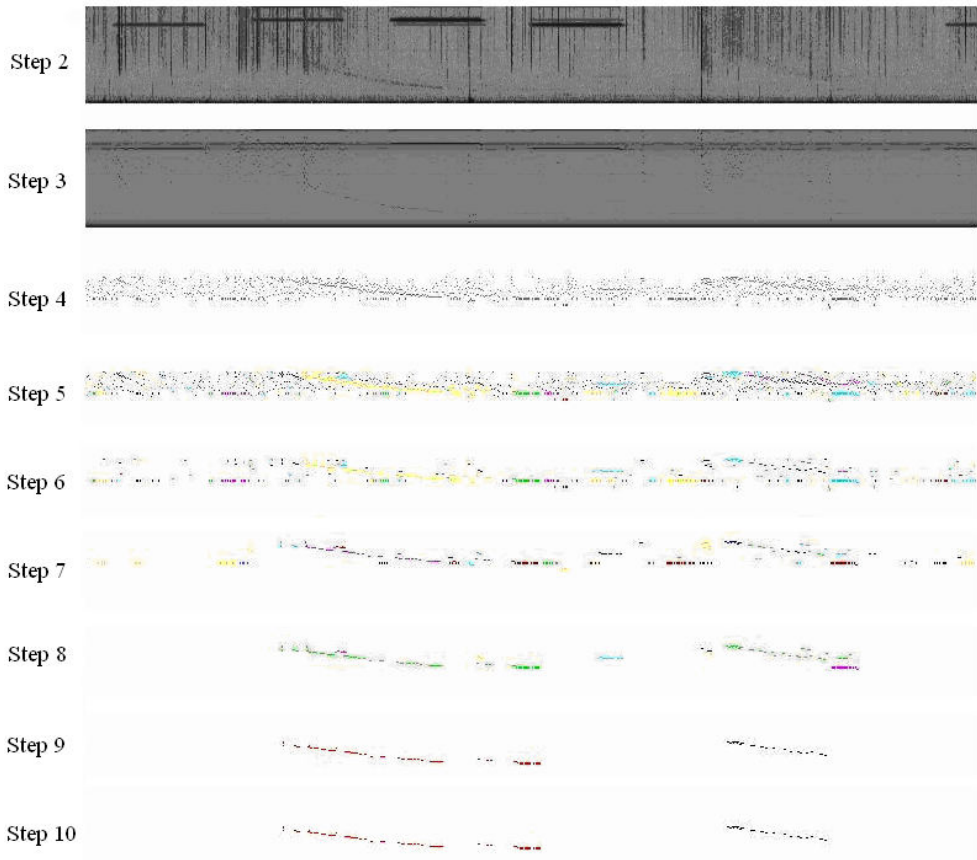B. Recursively join the groups having common points with each other.

Fig. 4. Examples of 2-10 steps.

Step 6. In each *group* ∈ *G*, remove the points that do not have neighbours vertically or horizontally (Figure 4 (step 6)).

Step 7. In each *group* ∈ *G*, calculate the number *Kdifx* of points with different values at the X-axis, and if *Kdifx*<P*difx*, then remove group *group* from *G* (Figure 4(step 7)).

Step 8. In each *group* ∈ *G*, using the method of least squares, calculate the approximating straight line y=k*x+b. Remove all groups *group*∈*G* having the slope coefficients *k>kmax* or *k<kmin*. Recursively join the groups considering parameters *maxDifX*, *kAbsDif*, interval *[difY1; difY2]* etc. (Figure 4(step 8)).

Step 9. Filter the joined groups (i.e. remove the groups that do not pass this filter) considering parameters *pJoinedMin*, *kminJoined*, *kmaxJoined*, *distJoined* etc. (Figure 4(step 9)).

Step 10. Using dynamic programming, calculate the maximal sum *SumLTRD* of the number of group points on the covered distance, allowing movement on group points right or down from the upper left corner of the rectangle into the lower right corner of the rectangle of the group, and calculate the maximal sum *SumRTLD* of the number of group points on the covered distance allowing movement on group points left or down from the upper right corner of the rectangle into the lower left corner of the rectangle of the group.

Calculate percentages *percentLTRD* and *percentRTLD* of the number of group points on the covered distance among the total number of group points. Using the tables *TLTRD* and *TRTLD*, group sizes and obtained values *percentLTRD* and *percentRTLD*, filter the groups.

The rest groups of points (logarithmic scale by frequency) are labeled as whistler points (Figure 4 (step 10)) and must be checked according to the functional scheme (Figure 2) by algorithms of detailed analysis and recognition of whistlers, as well as algorithms of the entire whistler's shape determination.

We should note the high speed of work of the proposed algorithm. For instance, a 15-minute WAV file with discretization frequency 44.1 kHz is analyzed in 7 seconds (including the time of reading the WAV file from the filesystem in RAM) on an Intel(R) Core(TM) i7-4702MQ CPU 2.20GHz processor. At the present time, the proposed algorithm is used on the stage of primary recognition of whistlers on mini-computer software-hardware complex at the radio physical observation stations Karymshino and Oybenkel, as well as recognition of whistlers is performed on the basis of open access data of abelian.org VLF station network in Todmorden United Kingdom (53,703N, 2.072W), Bielefeld Germany (52.146N, 8.458E), Cumiana Italy (44.96N, 7.42E), Warsaw Poland (52.16313N, 21.03094E), Australia Heathcote (36.804163S 144.67559E). The algorithm is launched with different sets of configured parameters, followed by results combination using the rules based on confidence coefficients. In figures 5-10, we provide some examples of whistlers recognized using the proposed algorithm in different location points.

Based on a large number of whistlers recognized at various location points, we created a training set of whistler images and images without whistlers (including mistakenly recognized whistler images). Based on this training set of images we with the help of the *Tesla K80* video card trained a deep neural network , which is based on the architecture of the convolution neural network *ResNet-50*. The neural network is currently used at the stage of detailed analysis of the whistler recognition results (Figure 2). The current trained neural network shows about 98% accuracy of classification on validation images, which not used at the learning stage. Two seconds duration test image of the spectrogram is recognized in 0.11 seconds on a computer without a video card with available two cores of the Intel (R) Xeon (R) CPU 2.30GHz processor.

Further research directions on the use of deep neural networks for the recognition of whistlers involves working on the selection of a suitable neural network architecture, the expansion of the training set, the use of a neural network to determine the whistler's shape entirely (not only its tail part), etc.
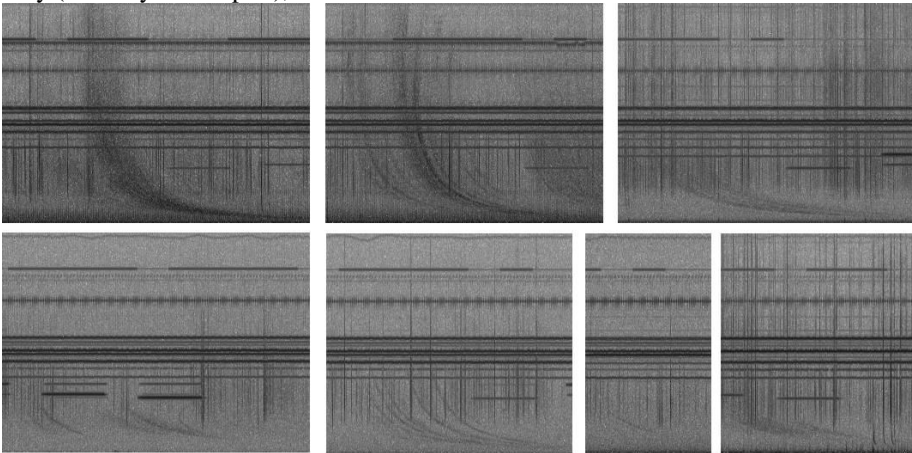

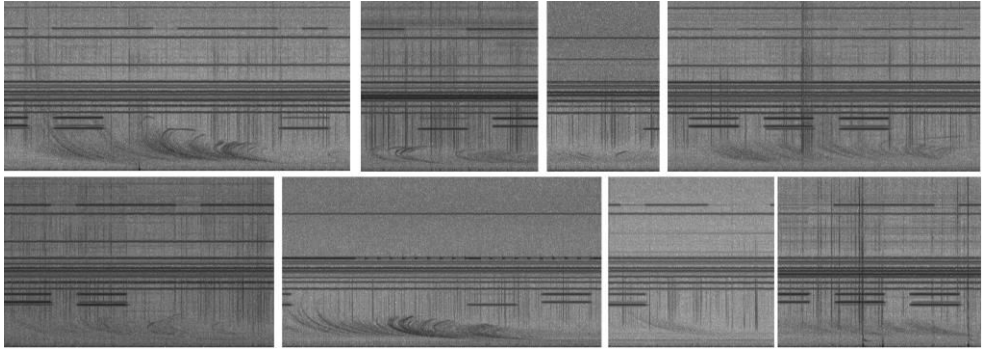**Fig. 5.** Examples of recognized whistlers at the Karymshino station in Kamchatka.

Fig. 6. Examples of recognized whistlers on Oybenkel Geophysical Observatory of the Yu.G. Shafer Institute of Cosmophysical Research and Aeronomy (the Sakha Republic (Yakutia)).
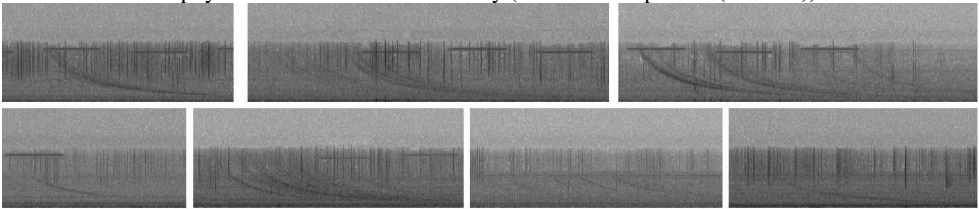
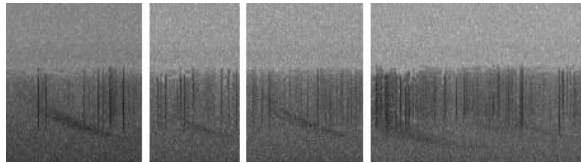

Fig. 7. Examples of recognized whistlers in Todmorden UK.

Fig. 8. Examples of recognized whistlers in Heathcote, Australia.
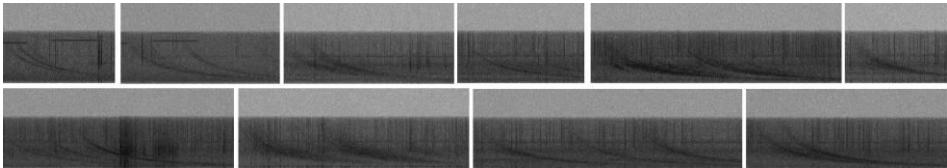

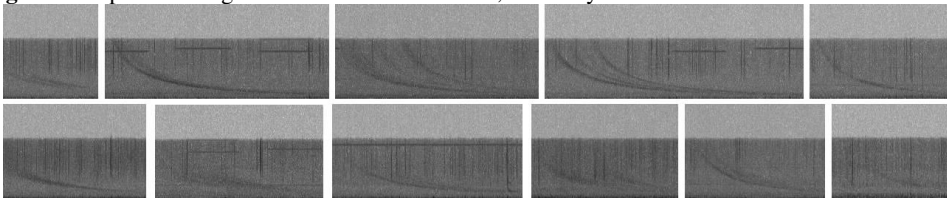**Fig. 9.** Examples of recognized whistlers in Bielefeld, Germany.

Fig. 10. Examples of recognized whistlers in Cumiana, Italy.
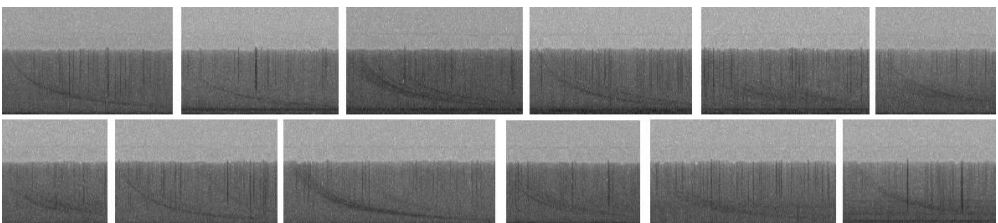

Fig. 11. Examples of recognized whistlers in Warsaw, Poland.

4 Conclusion

On the basis of the proposed algorithm, we created a historical database of whistlers registered with the help of a magnetic East-West antenna at the Karymshino station in Kamchatka and an automatically updated database of registered whistlers on mini-computer software and hardware complexes for distributed monitoring of low-frequency electromagnetic radiation [3]. The most correct and complete source WAV-files in the historical database are presented for 2012-2013 years.

Table 1. Spearman coefficients of correlation calculated for different dates.

Date	Spearman coefficients of correlation by the minutes of the day	Spearman coefficients of correlation by the hours of the day
10.02.2012	0.792	0.903
16.02.2012	0.868	0.947
17.02.2012	0.892	0.943
18.02.2012	0.774	0.869
19.02.2012	0.082	-0.044
20.02.2012	0.813	0.942
22.02.2012	0.593	0.603
14.03.2012	0.601	0.681
15.03.2012	0.654	0.831
16.03.2012	0.624	0.516
01.04.2012	0.595	0.682
08.04.2012	0.264	0.100
05.11.2012	0.357	0.317
17.11.2012	0.053	-0.082
24.12.2012	0.868	0.942
25.12.2012	0.922	0.873
26.01.2013	0.818	0.926
09.02.2013	0.689	0.724
10.02.2013	0.768	0.786
19.02.2013	0.405	0.445
20.02.2013	0.415	0.383
21.02.2013	0.641	0.609
22.02.2013	0.591	0.560
24.02.2013	0.776	0.801
26.02.2013	0.755	0.739
21.03.2013	0.101	-0.111
22.03.2013	0.449	0.620
24.03.2013	0.557	0.647
13.10.2013	0.008	0.027
22.10.2013	0.625	0.915

The results of the whistler recognition for 2012-2013 years allowed us to establish on the days of strong whistler activity, there is an average positive correlation between the number of whistlers registered at Karymshino station and the number of lightning strokes registered by WWLLN global network in the coordinate rectangle LAT 25S-45S, LON 140E-160E (Australia). Table 1 and some selected figures (12-13) shows the results of the Spearman correlation analysis described above by the minutes and hours of the day on days with the number of registered whistlers at the Karymshino station over 6,000.
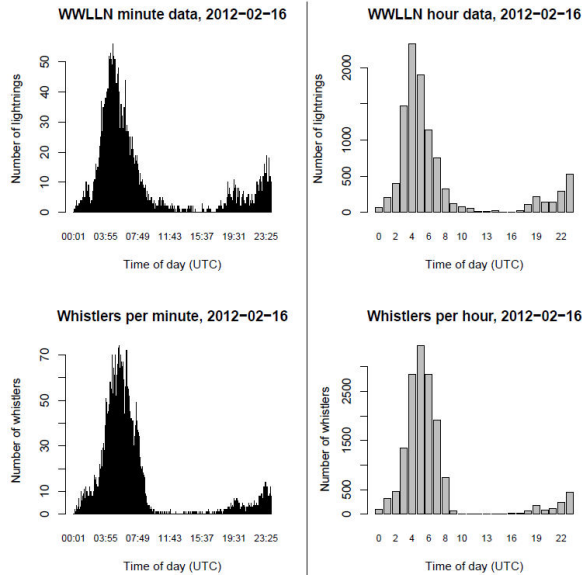

Fig. 12. Correlation between the number of whistlers registered at Karymshino station and the number of lightning strokes registered by WWLLN global network in the coordinate rectangle LAT 25S-45S, LON 140E-160E (Australia) on 2012-02-16.
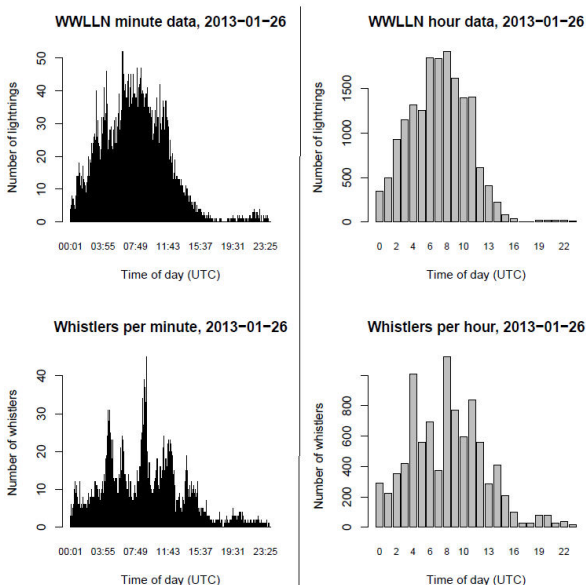

Fig. 13. Correlation between the number of whistlers registered at Karymshino station and the number of lightning strokes registered by WWLLN global network in the coordinate rectangle LAT 25S-45S, LON 140E-160E (Australia) on 2013-01-26.

References

1. V. A. Mochalov, A. V. Mochalova, Solar-Terrestrial Relations and Physics of Earthquake Precursors, E3S Web Conf. **20**, 11 (2017)

2. V. A. Mochalov, R. R. Karimov, G. I. Drugin, N. V. Cherneva, A. V. Mochalova, M. S. Permyakov, A. N. Droga, V. V. Argunov, A. A. VERSIM VLF/ELF Remote Sensing of the Ionosphere and Magnetosphere, 44 (2018)

3. V. A. Mochalov, G. I. Drugin, R. R. Karimov, B. M. Shevtsov, N. V. Cherneva, A. V. Mochalova, M. S. Permyakov, A. N. Droga VarSITI Newsletter, **16**, 8 (2018)

4. J. Lichtenberger, C. Ferencz, D. Hamar, P. Steinbach, C. J. Rodger, M. A. Clilverd, A. B. Collier Journal of Geophysical research, **115**, A12214 (2010)

5. VLF Receiver Toolkit — Notes,http://abelian.org/vlfrx-tools/notes.html#Whistler%20detection

6. E. A. Malysh, G. M. Vodinchar, N. V. Cherneva Vestnik KRAUNC. Fiz.-mat. nauki. **15:4**, 43 (2016)

7. K. S. Dharma, I. P. A. Bayupati, P. W. Buana Journal of Theoretical and Applied Information Technology, **66-2**, 638 (2014).