This paper presents a general approach to the improvement of speech intelligibility in broad band acoustical noise. By using the methods of Markov filtering the digital processing algorithms of noise-added speech are being synthesized and their experimental study is being carried out.

INTRODUCTION

The telephone communication systems and the systems of automatic man-machine communication by voice often operate in a severe broad band acoustical noise situations. The organising protective measures and the compensation techniques do not always provide the effective noise suppression. In such cases the signal-noise ratio (SNR) of the microphone output may be 0-3 dB, and the intelligibility S may be 40-50% /1,2/. The special digital processing for noise reduction is applied but it doesn't allow to increase intelligibility sufficiently so far /1,2/. The aim of this paper is to develop the effective processing methods by using Markov filtering.

FORMULATION OF THE PROBLEM

In interval the duration of which is about 20 - 50 ms the mixture of signal and noise is

\[ z_t = x(\lambda(t)) + n_t, t = 0, t_1, t_2, \ldots, \]

where \( z = (z_0, z_1, \ldots, z_m)^T \) is a vector of parameters describing the articulation apparatus state (ArA); \( x(\lambda(t)), n_t \) - the sample sequences of speech signals (SS) and noise. Because of the low accuracy of articulatory organs the parameters take a continuous set of values.

For the automatic recognition (reception) of corrupted speech the values of parameter \( \lambda \) should be classified. However, in the process of extraction it is quite enough while using \( z_t \) to formulate such signal \( u(\lambda^*, t) \) on hearing of which the maximum intelligibility is achieved:

\[ S_{z} \max = S(u), \]

where \( S(z) \), \( S(u) \) - is the intelligibility of signals \( z_t \) and \( u(\lambda^*, t) \) respectively. Since a vector \( \lambda^* \) or an unknown function \( g(\lambda) \) is classified in the process of human perception, the value of \( \lambda^* \) should be chosen in such a way that \( E_\lambda = E[\lambda^* - \lambda]^T Q_\lambda [\lambda^* - \lambda] = \min \)

where \( E \) - mathematical expectation operation, \( Q_\lambda \) - a weighted coefficient matrix. The minimum attainable value \( E_\lambda \) is defined by Kramer-Rao's inequality.

So the problem of speech extraction is interpreted as the construction of the
The function $p$ is found on the synthesizer, it may be the "Mr 31 clipping function. For example, $x(t)$ may be the clipping of $x(t)$ to the articulatory tables of syllables without any pauses are shown. The probability of a classification error of tone-consonant-pause is about $10^{-2}$ with a zero threshold.

**THE FILTERING ALGORITHMS**

The filtering algorithms can be used to find a family of curves with the equal reception accuracy: $\text{SNR} = \frac{\text{Signal}}{\text{Noise}}$, with parameter $x(t)$. The result is shown in fig. 2.

![Fig. 1. A General Diagram of Speech Extraction](image1)

**SIGNAL AND NOISE MODELS**

For the best evaluation of $\lambda$ the adequate models of signal and noise are required. The simplest model of the broadband noise is Gaussian sequence $\eta(t)$ with $E\{\eta(t)\} = 0$, $E\{\eta(t)\eta(t')\} = \delta(t-t')$. The more precise model is the process of autoregression

$$y(t) = \sum_{i=1}^{m} a_i x(t-i) + \xi(t), \quad x(t) = \sum_{i=1}^{m} c_i y(t-i) + \xi(t),$$

where $a_i$ is evaluated a priori by the noise realization by means of a least-square technique with limitations. The orthogonal functions $y_i$ may be found experimentally. To do this on the speech signal of a concrete speaker the set of the parameters values is defined in models (3), (4), then a set of autoregression functions is given $y(t-1),\ldots, y(t-m)$, and the Karhunen-Loeve basis is built for it.

![Fig. 2. The Function of Excitation](image2)

**THE EVALUATION OF THE POSSIBLE INTELLIGIBILITY**

The intelligibility $\lambda$ may be evaluated in the presence of noise with an average flat spectrum. Consider $z_0(t) = x(\lambda,t) + n_0(t)$, $z_1(t) = f[z(\lambda,t)] + n_1(t)$.

Choose the function $f$ so that $\text{SNR} = \frac{\text{Signal}}{\text{Noise}}$. For example, $f$ may be the central clipping function, and the threshold of the clipping $x_{th}$. Putting down the Kramer-Kahn's inequalities the formulas for $\lambda_1$ (SNR) and $\lambda_\text{max}$ (SNR) may be obtained. In the situation where $\lambda_1$ (SNR) is

$$\lambda_\text{max} (\text{SNR})$$

we can find a family of curves with the equal reception accuracy: $\text{SNR} = \frac{\text{Signal}}{\text{Noise}}$, with parameter $x(t)$. The result is shown in fig. 2.

![Fig. 3. Signal spectrum $G_x, G_y, G_z$](image3)

**EXPERIMENTAL RESULTS**

Testing of algorithms No. 1-3 are performed on the speech signal with the sampling frequency 15 kHz and with the number of quantizing levels $2^{12}$. In fig. 3 the power spectral densities of the initial $(G_x)$, the processed $(G_y)$ and the noise-added speech $(G_z)$ are shown for the word "agreement" (algorithms No. 1, 2). In fig. 4 the curves of likelihood function $A$ and the current signal power $E_x^2$ received on the articulatory tables of syllables are shown. The probability of a classification error of tone-consonant-pause is about $10^{-2}$ with a zero threshold. In Table No. 1 the results of tests are shown, where $\Delta \text{SNR}$ is a benefit of the algorithms.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>No. 1</th>
<th>No. 2</th>
<th>No. 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta \text{SNR}$</td>
<td>7</td>
<td>6</td>
<td>10</td>
</tr>
</tbody>
</table>

![Fig. 4. Likelihood Functions $\Lambda$](image4)
In $SNR$, the number of algorithm 3-1 is a sequential application of the algorithms $N3$ and $N1$. These results are achieved for the noise with close to an average flat spectrum and model (3), (4).

In pauses the mixture of $x_t$ is multiplied to a coefficient $q < 1$. The coefficient of noise power in filters is chosen experimentally.

In Table No.2 the signal-error prediction ratio ($SER$) for models (4), (5) which is achieved on the initial speech signal is given. There the results of algorithm No. 2 with (5) in noisy environment because of the engines operation.

In Table $\sum$ is percentage of the real favours given by the listeners to the processed signal. The number of listeners is 20-25.

<table>
<thead>
<tr>
<th>Model</th>
<th>$SER$, dB, for N</th>
<th>$\sum$</th>
<th>ASNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>(4)</td>
<td>21, 24, 26, 27, 27, 27</td>
<td>85-88</td>
<td>3-5</td>
</tr>
</tbody>
</table>

CONCLUSION

The method of intelligibility improvement in noisy environment is worked out. The theoretical benefit of the digital processing for noise with long-term average flat spectrum is evaluated. By using the Markov filtering techniques the algorithms of mutual speech filtering, the parameters evaluation and the classification of tone-consonant-pause are developed. The algorithms provide the improvement of the corrupted speech intelligibility in broad band noise and can be technically done on the mikroprocessor devices Am 2900.