The European Proceedings of
Social & Behavioural Sciences
EpSBS

Future Academy ISSN: 2357-1330

http://dx.doi.org/10.15405/epsbs.2017.12.12

ICPE 2017
International Conference on Psychology and Education

THE EXPERIMENTAL RESEARCH OF THE WAY ACOUSTIC NOISE INFLUENCES SPEECH CHARACTERISTICS

A. Sh. Gabdrakhmanov (a), O. N. Korsun (b)*
*Corresponding author

(a) Moscow Institute of Physics and Technology (State University), Moscow, Russia, besha5500@yandex.ru
(b) State Research Institute of Aviation Systems State Scientific Center of Russian Federation, Moscow, Russia, marmotto@rambler.ru

Abstract

One of the promising direction of the development of interface cockpits of modern aircraft is the voice control of on-board equipment. The main element of the voice control is the automatic recognition of voice commands, the effectiveness of which is determined by many factors, first of all the different types of acoustic noises. Noise is one of the biggest problems in speech recognition area. It does not only create problems when finding a speech signal in condition of noise or another acoustic interference, but also can change the speaker`s voice manner. This article focuses on the results of the way acoustic noise influences speech characteristics. The authors have come up with the two methods to decrease this influence. Firstly, recognition patterns can be formed using words, which have been influenced in this way. Secondly, the special reverse, or feed-back, channel can be added in the speech recognition system, which enables a speaker to hear himself as well as usually. All results were obtained with the implementation of the speech recognition algorithm, based on parameterization of the speech signal, on method of forming a pattern and the method of recognizing the individual words by comparison with standards of words, on linear regression and application of an additional microphone. An additional microphone is positioned at a distance of 0.3–0.7 m from the speaker in order to fix predominantly acoustic noises in the cockpit.

Keywords: Acoustic noise influence, speech recognition, noise resistant algorithm, speech characteristics

http://dx.doi.org/10.15405/epsbs.2017.12.12
Corresponding Author: O. N. Korsun
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

1. Introduction

Resistance to the acoustic noise is one of the most important requirements for the voice control system of on-board equipment in the cockpit, where the noise maximum level can amount 80–90 dB.

The main aim of this study is to carry out the experimental research of the way acoustic noise influences a speaker`s voice characteristics, which decreases the efficiency of the recognition algorithm developed (Korsun, 2017). The percentage of correctly recognized words for a small dictionary system has been chosen as a criterion.

Before the recognition all speech signals should be transformed into vectors of parameters, which are obtained by taking a Fourier transform of a short time window of speech (frame) and calculating the spectrum in one out of the three standard ways (Rabiner, 1993): spectrum, mel-spectrum, mel-cepstrum (Randall, 2016, Wisniewski, 2008). Only after such a transformation they can be further compared with the patterns from the dictionary. The pattern in this case is the average parametric vector of the series of the word implementations of one or more speakers.

Firstly, in case of silence our algorithm recognizes correctly 99% of words (Rabiner,1993). This level has been chosen as a required one. It is the highest level in known algorithms (Chia, 2012, Gong, 1995, Pal, 2015, Gonzalez Jose, 2016).

The next stage of the experimental research presupposes the program "mix" of the words records and acoustic noise record. This technique ensures a strict compliance with the assumptions of the noise additivity and excludes the noise influence on the speaker`s speech. The initial algorithm shows, as expected, the increased number of errors caused by the noise. The further application of the modified algorithm (Korsun, 2017), based on the additional microphone and multiple linear regression, once again enabled to correctly recognize words with the required 99% probability. However, the application of this algorithm in terms of real noise brought no positive results. That proves our hypothesis that the noise leads to changes in speech characteristics.

2. Problem Statement

The paper is aimed at carrying out the experimental research of the way acoustic noise influences a speaker`s voice characteristics.

3. Research Questions

The study had the following objectives: first, to examine the way acoustic noise influences a speaker`s voice characteristics. Secondly, to develop the method of decreasing such influence.

4. Purpose of the Study

The main goal of the present study is to find out the way we could reduce the influence of noise on the speech characteristics, which decreases the efficiency of the recognition algorithm developed.

http://dx.doi.org/10.15405/epsbs.2017.12.12
Corresponding Author: O. N. Korsun
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

5. Research Methods

The percentage of correctly recognized words for a small dictionary system has been chosen as a criterion in both methods. Methods are listed below.

5.1. The noise in the headphones

The experiment with the noise in the headphones (see Fig. 1) was conducted to test the hypothesis about the way noise influences the speaker`s speech characteristics (Wisniewski, 2008). By the noise in the headphones we mean the record of the real noise in the cockpit of the aircraft Boeing 747 was fed to the speaker's headphones while the words were pronounced, so the noise did influence the speaker but the speech registration included no noise. The experiment consisted of the three stages. Firstly, in headphones there was no noise. Secondly, the record of the noise with loudness of 80 dB was played. Thirdly, volume was set to 90 dB. The volume level was measured by sound level meter. At the two last stages the speaker almost ceased to hear himself. Therefore, a natural reaction to such an external condition was an attempt to anguishly speak up.

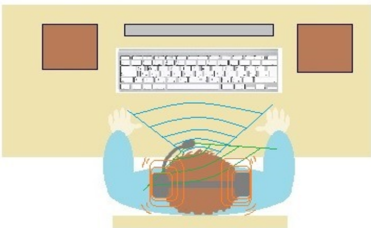

Figure 01. Scheme of the experiment with noise in the headphones.

To recognize all the words from this experiment we used the last version of the algorithm with the multiple linear regression.

The pattern "N-v 0 dB" was formed on the words without noise in the headphones from the first stage using only the speaker`s N-v records. Also were formed serial of patterns for each of the speakers and for each of the two amplitudes of acoustic interference "80 dB" and "90 dB".

5.2. A reverse channel

In addition to the use of the right patterns to reduce this influence we recommend the use of a special reverse channel (Gong, 1995), or an acoustic feed-back. This reverse channel enables to play the signal through the headphones at real time from the headphones' microphone. In this experiment, even under the influence of a strong acoustic noise produced by a stereo system, which was being played in the room while recording, the speech characteristics underwent minimal changes. This effect is achieved due to the fact that the speaker hears his own speech and does not distort his voice, as in the experiment before. The scheme of the experiment is presented in figure 7.

http://dx.doi.org/10.15405/epsbs.2017.12.12
Corresponding Author: O. N. Korsun
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

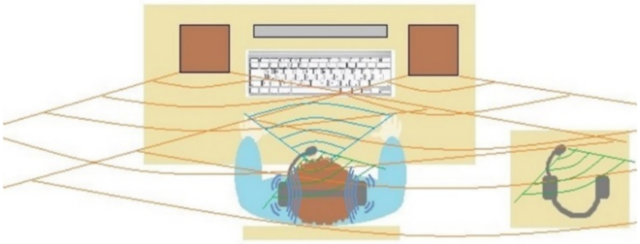

Figure 02. Scheme of the experiment with reverse channel.

6. Findings

The results of the special experiments are described in 2 parts.

6.1. Experiment with the noise in the headphones.

Figure 2 shows the recognition results from the first phase of the experiment, in which the noise in the headphones was off. It is seen that all three parameterization types show a good quality of recognition. The number of errors does not exceed 1% of the total number of words for all speakers. For the pattern "N-v 0 dB" the percentage of the total number of errors for all 4 speakers was 0% for spectrum (spectr) and mel-spectrum (mel) methods of parametrization and only 0.6% for the mel-cepstrum (cepstr) method.

It should be noted that patterns formed in the in conditions of the headphone noise with levels "80 dB" and "90 dB" ("N-v 80 dB" and "N-v 90 dB") did show some increase of errors but the maximum error level didn't exceed 1%.

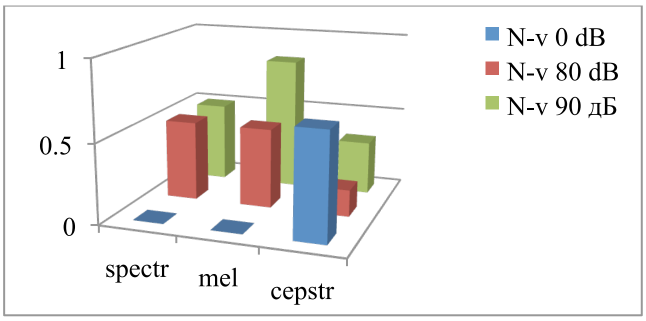

Figure 03. Total number of errors in word recognition without noise.

However, for words with the noise of 80 dB in the headphones, pattern "N-v 0 dB" ensures the recognition with the increase in the total number of errors for all 4 speakers to 4% in case of the spectrum method of parameterization, to 3.2% for the mel-spectrum method and to 1% for the mel-cepstrum method (Fig.3).

http://dx.doi.org/10.15405/epsbs.2017.12.12
Corresponding Author: O. N. Korsun
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

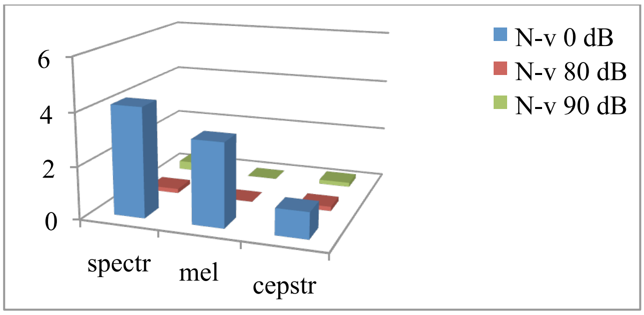

Figure 04. Total number of errors in word recognition with the noise in the headphones 80 dB.

Figure 3 also shows that pattern "N-v 80 dB" formed on the N-v`s records, but on the words with the noise of 80 dB, enable us to recognize the words with the noise 80 dB in the headphones much better. The total number of errors is 0.2% for the spectrum, 0% for the mel-spectrum and 0.2% for the mel-cepstrum. Using the pattern "N-v 90 dB" we observed an insignificant increase in the number of errors to a level of 0.3% for the spectrum type.

For the words with the noise of 90 dB the total number of errors for the pattern "N-v 0 dB" is respectively 4.2% for the spectrum, 3.7% for the mel-spectrum, 0.5% for the mel-cepstrum (Fig. 4).

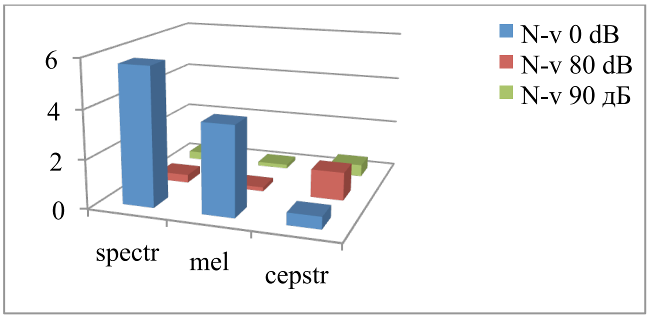

Figure 05. Total number of errors in word recognition with the noise in the headphones 90 dB.

Pattern "N-v 80 dB" for the records with the noise of 90 dB shows the results– 0.4% for the spectrum, 0.2% for the mel-spectrum and 1.2% for the mel-cepstrum. The pattern "N-v 90 dB" in comparison with the "N-v 80 dB" reduced errors by 0.7% for the mel-cepstrum parameterization method, while the level of errors for other methods were not changed.

Figures 5 and 6 describe how the patterns, formed from words with the noise for each of the speakers, improved the recognition in comparison with the pattern "0 dB".

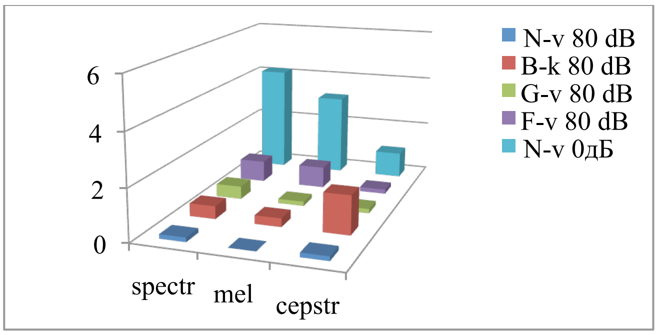

Figure 06. Total number of errors for the words with the noise in the headphones 80 dB.

http://dx.doi.org/10.15405/epsbs.2017.12.12
Corresponding Author: O. N. Korsun
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

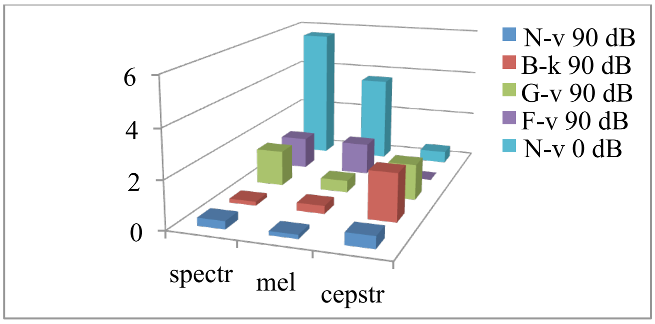

Figure 07. Total number of errors for the words with the noise in the headphones 90 dB.

Thus, the right choice of a suitable pattern enables us to recognize words with high efficiency, even despite the influence of noise on the speech parameters of the speaker. The patterns with noise shows a good level of recognition not only in the recognition of words with noise, but also in the recognition without noise, for which there are no changes to voice characteristics of the speakers under the loud acoustic noise.

6.2. The use of a reverse channel.

Table 1 shows that the use of a second microphone in combination with the choice of a suitable model and the type of parameterization can give the required detection level (number of errors less than 1%). So for the really presented in the room while recording two types of noise "Speech" and "Boeing" pattern "N-v 80 dB" minimizes the total error for all four speakers to the value of 0.8% and spectrum type of parameterization.

Table 01. Summary results of word recognition using reverse channel, %

Mic - Pattern	0 dB			Speech 80 dB			Boeing 80 dB		
	sp	mel	cep	sp	mel	cep	sp	mel	cep
1 mic – N-v 0 dB	0.3	0.3	0	6.3	4.7	3.8	0.7	0.3	21.7
1 mic – N-v 80 dB	1.5	5	0.3	5.3	3.3	7.5	0.8	2.5	34.2
2 mic – N-v 0 dB	0.3	0.3	0	2.5	3.8	2.8	0.5	0.7	7
2 mic – N-v 80 dB	1.5	5	0.3	0.8	1.7	3.7	1.5	3	1.2

It is recommended to use the pattern "N-v 0 dB" for the noise "Boeing 80 dB", as it gives us the total number of errors equal to 0.5 % and 0.7% for spectrum (sp) and mel-spectrum (mel) types while using both microphones.

The efficiency of the reverse channel application is illustrated in figure 8, which shows the total number of errors for the two types of noise using an additional microphone and the pattern "N-v 0 dB".

http://dx.doi.org/10.15405/epsbs.2017.12.12
Corresponding Author: O. N. Korsun
Selection and peer-review under responsibility of the Organizing Committee of the conference
eISSN: 2357-1330

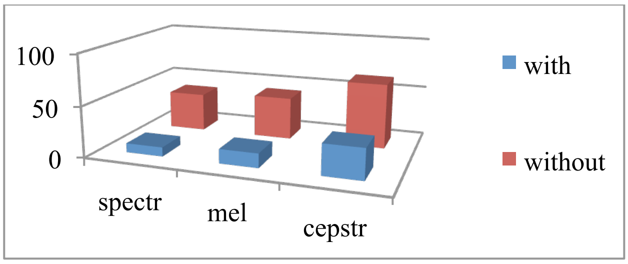

Figure 08. Comparison of the number of errors with and without the use of the reverse channel. Standard "N-v 0 dB". The information from the two microphones has been analyzed.

Thus, the use of the reverse channel reduces the number of errors for records with real noises. In addition to the reverse channel you must still take into account the correct choice of the patterns, the effectiveness of which depends on each specific casePlease replace this text with context of your paper.

7. Conclusion

The results of the special experiment (noise of 80-90 dB is supplied only to speaker`s headphones) showed that the speech characteristics of the speaker are changed significantly. However, the percentage of recognition errors using patterns, formed from the records without noise, is 3.2...4.2% for the spectrum and mel-spectrum variants of parameterization. For patterns, formed using words with noise of 80-90 dB, the percentage of recognition errors is reduced and is 0...0,4% for the spectrum and mel-spectrum variants of parameterization.

To reduce the influence of noise on the speech characteristics the reverse channel is proposed. The signal from the main microphone is led to the headphones, so that the speaker could hear himself in the conditions of strong noises as well as in normal conditions.

The experiment on speech recognition in real noise conditions and using the reverse channel were conducted for 4 speakers and absolute noise level of 80 dB for noise types "Boeing" and "Speech". The results of the experiment showed that using the patterns, generated on the data without noise, we can obtain the minimum error levels for the spectrum variant of the parameterization and the range 0.5...2.5%, that 1.5...3 times less than the same case without the use of the reverse channel.

Acknowledgments

This research is supported by Russian Fund for Basic Researches (RFBR), project 15-08-06946.

References

Chia Ai O., Hariharan M., Yaacob S., Sin Chee L. (2012). Classification of speech dysfluencies with MFCC and LPCC features. *Expert Systems with Applications*, 39, 2157–2165.

Gong, Y. (1995). Speech recognition in noisy environments: a survey. *Speech Communication*, 16(3), 261-291.

Gonzalez Jose A., Gomez Angel M., Peinado, Antonio M. (2016). Spectral Reconstruction and Noise Model Estimation Based on a Masking Model for Noise Robust Speech Recognition. *Circuits Systems And Signal Processing*, 36(9), 3731-3760.

Korsun O. N., Gabdrakhmanov A. Sh. (2017). Speech recognition based on relations with all the patterns in the dictionary. *Vestnik komp`iuternykh i informatsionnykh tekhnologii*, 1, 10-15.

Pal M., Saha G. (2015). On robustness of speech based biometric systems against voice conversion attack. *Applied Soft Computing*, 30, 214-228.

Rabiner L., Juang B.H. (1993). *Fundamentals of Speech Recognition*. Prentice-Hall International, Inc., Englewood Cliffs, New Jersey, 507.

Randall, Robert B. (2016). A history of cepstrum analysis and its application to mechanical problems. *Mechanical Systems and Signal Processing*, 97, 3-19.

Wisniewski M., Kuniszyk-J´o´zkowiak W., Smolka E., Suszynski W. (2008). Automatic detection of disorders in a continuous speech with the hidden Markov models approach. *Proceedings of Computer Recognition Systems 2*, 45, 445–453.