Developing an algorithm for creating the optimal spectrum of a masking noise considering a speaker's diction

A V Ivanov, E E Shemshetdinova

Novosibirsk State Technical University, Novosibirsk, Russia

Abstract. The experimental research is carried out to build up a database with speaker's voices. The structure of the database with speaker's voice signals is designed. The software to calculate the optimal envelope shape of interference based on the speech samples is developed. The obtained interference is suitable to disguise voice signals of certain speakers in protected premises during the confidential conversation. Development prospects are described, and recommendation for further research is given.

1. Introduction

Nowadays, an approach based on the formant method of N. B. Pokrovsky [1, 2] is applied to calculate word intelligibility and to estimate the protection level of voice information.

The approach has been developed for conditions that significantly differ from information security tasks (for estimating the quality of communication lines). That has led to a number of shortcomings discussed by many authors in order to adapt the approach to solve information security problems [3-7]. The approach considers using the A-weighted levels of voice signals measured at the distance of 1 m from the speech source, i. e. an averaged voice spectrum. However, a voice spectrum is known to differ for various people.

The research on influence of variety in voice spectrums on the result of speech intelligibility estimation has been conducted to understand if use of the averaged voice spectrum is possible in problems of speech intelligibility estimation. The findings have shown that a speaker's voice spectrum affects greatly the result of speech intelligibility estimation [8].

As a result, the following recommendations have been given - if the number of speakers discussing information in a protected room is limited, the protection level should be estimated based on their voice spectra. If following a recommendation is not possible, the protection must be estimated according to the voice spectrum providing with the higher speech intelligibility level instead of the averaged voice spectrum.

Considering this effect can significantly increase the protection level of voice secret and confidential information processed in premises.

In this regard, we have made a decision to develop a software to automatise solving the problem on finding the optimal interference for a certain speaker or a group of speakers in premises with secret or confidential information and calculating all the required parameters for information security equipment.

2. The experimental design

To achieve the goal, the following tasks must be solved:

1. To study the standard approach to voice information estimation;
2. Conduct experimental research to build the database up;

3.   Develop an algorithm and a software to obtain the parameters for setting information security equipment;

4.   Test the system.

3. Theory

1.   Review of modern state of the art

To protect voice information means to create interference in envelope of protected premises. For this goal, active information security equipment is applied. Interference is generated by a generator and radiated by transducers fixed on room walls. Places of possible information retrieval must be provided with the given signal-to-noise ratio, and white noise is usually used as interference.

Attention must be paid to setting active defence means to guarantee voice information is protected from intercepting.

In most cases, a proper attention is not paid to setting the means, and, to simplify the task, the highest level of A-weighted noise in octave bands is set instead of required. Such redundancy is not comfortable for people in the room.

It can lead to negligent performance of their duties, ignoring the instruction with requirements of using equipment for voice information security or to unauthorized operation of the equipment.

Switched on or incorrectly configured protection equipment creates an information leakage channel, and as a result creates a threat to information security.

Thus, influence of equipment for voice information security on people working in premises is still a problem.

Therefore, minimizing the parasitic noise level is a task of a proper configuring active protection equipment [9].

2.   A review of the common approach to speech intelligibility estimation

The modified formant approach is applied for estimating voice information protection. The most popular modification is Pokrovsky approach to word speech intelligibility estimation. The approach [1, 2] is based on measuring the acoustic pressure level (vibratory acceleration) of the following signals in octave bands:

- a test signal generated by applying a white-noise generator or a set of tone signals with central octave band frequencies;

- ambient noise or noise generated by active protection equipment and measured behind the envelope with a test signal generator switched off and active protection equipment switched on (if available);

- the signal-plus-noise mixture measured behind the envelope with a test signal generator switched on. The speech intelligibility level is calculated based on the obtained results.

The shortcoming of the approach is applying the averaged voice spectrum. It is inappropriate since voice spectra of speakers influence on the resulting speech intelligibility.

4. Recording speaker's voices to fill the database in

Eleven speakers (7 men and 4 women) with no speech disorders and special features have participated in the experiment on voice recording and estimating the level of acoustic pressure.

The speakers have read aloud an excerpt of a technical text during a minute. Technical texts are close to natural speech than articulatory tables (a special data set of the common sound combinations in a natural language).

The noise level meter ZET 110 and the BC501 microphone have been used to make records. The microphone frequency was set to 1 kHz by the SVAN SV35 calibration device. The record was conducted with the sampling rate of 50 kHz.

To increase the accuracy, the resulting levels for all octave bands and each speaker have been calculated by averaging of two records.

After the audio files being recorded, the following limitations have been designed:

- Audio files are recorded in the same conditions:
  - The duration is the same (one minute);

o   The same equipment is applied to record the speakers;
o   The distance between a speaker and the recording device is the same (1 meter);
• The recorded audio files are in .wav format;
• Space characters are not used in file names.
• The audio files must be recorded as monophonic sound (unlike to stereo, the sound is recorded and sent via the only channel).

5. The software design
1.   User requirements
The scenarios of applying the software are described. The scenarios assume that a user needs to adjust the parameters of information security equipment to the certain speaks staying in the attested premises currently.
1.   In user interface, a user forms a group of people staying currently in the room. For that:
a.   A user selects speakers from the list of the added in the database before;
b.   A user records speakers not added to the database following the limitations described in section 7.1.4 and uploads a received audio files to the system.
2. A user selects the word speech intelligibility level required for the room:
a.   A detailed description of the conversation content cannot be made – 0.6;
b.   A brief description of the conversation content cannot be made – 0.4;
c.   A conversation subject cannot be known – 0.3;
d   Conversation progress cannot be confirmed – 0.1 [4].
3.   After click on the "Calculation" button, a user receives the information on the A-weighted noise level must be applied to information security equipment.

2.   Functional requirements
The following requirements are imposed to the software:
•   Based on the input data (audio files with speaker's voice records), the voice spectra are obtained for further calculation of A-weighted levels in octave bands.
•   Based on the obtained frequency and corresponding spectrum level, the integral levels for octave bands are calculated and then normalized so the total level is equal to 70 dB as to the averaged spectrum.
•   The averaged voice spectrum is applied to the approach [1, 2] as the test signal levels; the spectrum of formant-like interference is used as noise. The speech intelligibility (W) levels are calculated according to the given signal-to-noise ratios q.
•   According to the needed for a user value W, the software selects the most appropriate of the obtained pair of values.
•   After obtaining the value W and corresponding q, q is applied to the sensation level formula Q_i and seven values of Q_i for each octave band are found.
•   Parallel to the calculations described above, we are looking for the "worst" speaker to protect conversation based on his intelligibility level that is the highest among the speakers. First, we calculate speech intelligibility W for each speaker according to [1, 2] and look for the result higher than others.
•   The formula to calculate Q_i is transformed into an equation with the unknown L_N_i. The value of the "worst" speaker is substituted into equation L_speech. The equation with one unknown, we obtain the integral noise level for each octave band.

3.   Choosing IDE and a programming language
C#, HTML5, JavaScript have been used for developing the site.

4.   Determining the structure of the database with speaker's voices
MySQL has been used in the project. The database consists of one table containing the paths to file storage.

Table 1. Files.

Field	Field type	Descrption
ID	int	Record ID
Name	varchar	WAV-file name
Path	varchar	Path to WAV-file on server
Speech_FileName	varchar	Path to integral voice levels on octave bands (DAT-files)
W_Path	varchar	Path to file with intelligibility for each noise type (DAT-files)

5.   The software architecture design

The software architecture is shown in fig. 1:
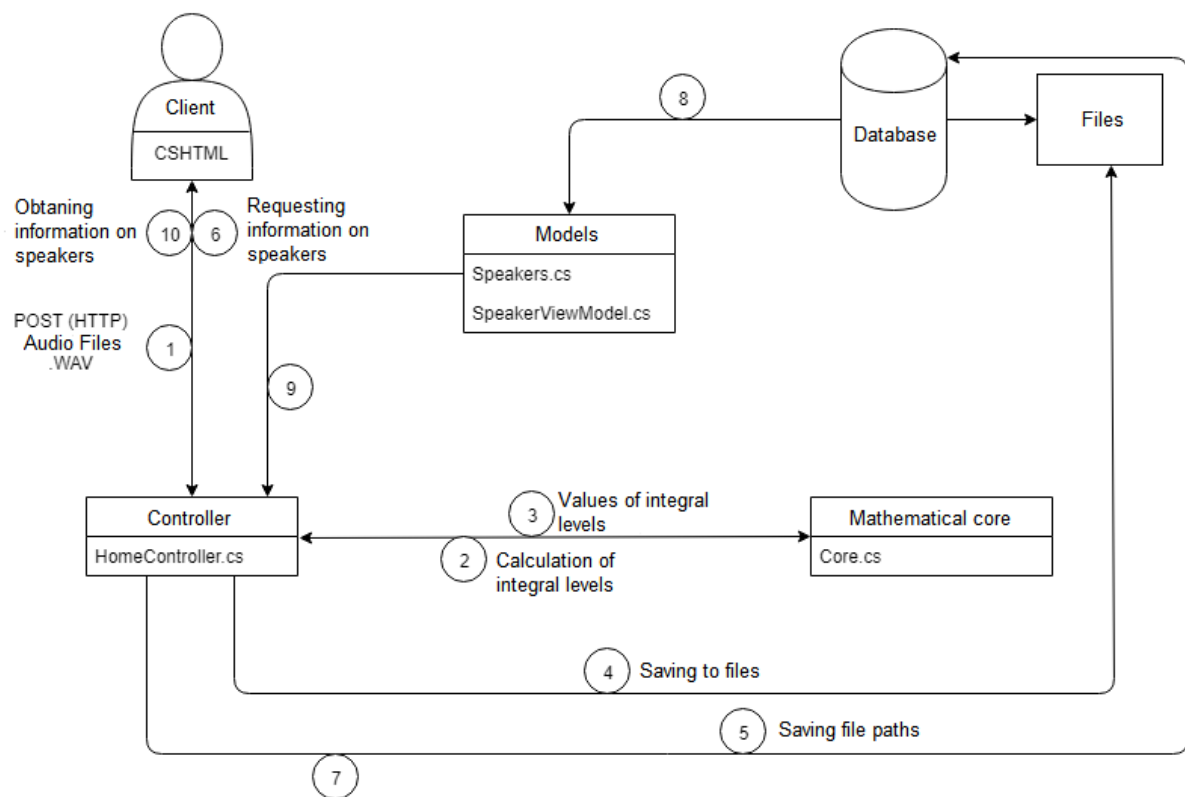
Figure 1. The software architecture.

The developed software with the "client-server" architecture is presented in fig. 2:

Figure 2. The main page.

The interface allows users to:
- upload an audio file,
- select the word speech intelligibility level required for the given premises,
- obtain information on the integral noise level to set it up for information security equipment.

6. Algorithm testing

To test the algorithm, a speaker with the highest speech intelligibility level after adding formant-like interference to the voice spectrum has been chosen. The signal-to-noise ratio equal to -10 dB has been selected. For this ratio, formant-like interference, and the averaged voice spectrum of Russian speech, word speech intelligibility is equal to 0.2 (20%). The values of speech intelligibility range from 0.07 (7) up to 0.28 (28%) for the whole group of speakers.

At the next step, optimal interference considering the voice spectrum of the speakers has been calculated for each of them. With the same signal-to-noise ratio (-10 dB), the values of speech intelligibility range between 0.08 (8%) and 0.25 (25%) for the whole group of speakers.

It should be noted that the biggest difference in signal-to-noise ratios compared for two interference is equal to 5 dB (male speakers #6 and #4). The speaker with the smallest intelligibility level has a difference between signal-to-noise ratios of two applied interference equal to 3 dB. Thus, taking only the spectrum of a speaker with the initially worst result into account is not reliable.

Obviously, as the original signal-to-noise ratio (and thus the speech intelligibility level) is growing, "benefit" of applying the interference calculated by the suggested algorithm increases too.

7. Conclusion

State of the art review and experiment conduction have shown a shortcoming in the conventional approach to voice information estimation. The approach is based on using the averaged voice spectrum. Since a speaker's voice spectrum significantly influences on the result of speech intelligibility, applying the averaged spectrum is inappropriate.

The protection level is recommended to be estimated based on the voice spectra of the specific speakers participating in the conversation instead of the averaged voice spectrum.

The approach to solving a problem of reducing the A-weighted noise level of equipment for active protection of voice information is suggested. The approach is based on applying the developed algorithm for calculating optimal masking noise for certain speakers.

The software based on the suggested approach is developed. Depending on the speaker, the calculated optimal interference with equal W is checked experimentally to be quieter by 3-5 dB than the formant-like interference.

The developed software can be employed for attestation of premises and initial configuration of security equipment as well as for adjustment of the configuration when speakers have been changed in the room.

In the future, the experiment will be repeated with applying a subjective method of speech intelligibility estimation (involving a group of speakers and auditors) to confirm the results. The current results are based solely on the calculation approach [1, 2].

References

[1]   Zheleznyak V K, Makarov Yu K and Khorev A A 2000 "Some methodological approaches to evaluating the effectiveness of information security speech," in Special Equipment, vol. 4 pp 39–45

[2]   Pokrovsky N B   1962 "Calculation and measurement of speech intelligibility," Moscow: Svyazizdat

[3]   Avdeev V B   2013 "About some directions of improvement of the methodical approaches applied at an assessment of efficiency of technical information security," in Special Equipment, vol. 2 pp 1–10

[4]   Didkovskii V S, Didkovskii M L and Prodeus A N 2008 Acoustic speech communication channel expertise," Kiev: Imeks Publ.

[5]   Gerasimenko V G, Lavruhin Yu N, and Tupota V I 2008 "Methods of protection of the acoustic speech information leakage via technical channels," Moscow: RCIB «Fakel»

[6]   Ivanov A V, Reva I L and Truschin V A   2010 "Experimental assessment of intelligibility of speech in problems of information security on the basis of the modified articulation measurements," *2010 X International Scientific-Technical Conference on Actual Problems of Electronics Instrument Engineering (APEIE),* Novosibirsk pp 133–136

[7]   Khorev A A 2010 "Monitoring the effectiveness of the protection of the premises allocated by the leakage of voice information through technical channels Information," in Security. Inside n. 1, pp 34–45

[8]   Ivanov A V Investigation of influence of speech specter differences on intelligibility estimation results / A. V. Ivanov, I. L. Reva, E. E. Shemshetdinova // Journal of Physics: Conference Series. 2018 Vol. 944 : Applied mechanics and dynamics systems : 11 intern. sci. and technical conf. , Omsk, 2017. Art. 012044 (5 p.). DOI: 10.1088/1742-6596/944/1/012044.

[9]   Ivanov A V 2018 Estimating protection of information from leakage through vibroacoustic channels: study guide // Novosibirsk: NSTU Publ. 73 p