ISSN 2278-3091
Volume 8, No.6, November – December 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse58862019.pdf
https://doi.org/10.30534/ijatcse/2019/58862019

Automation of Speech Quality Assessment in Speech Rehabilitation

Dariya Novokhrestova[1], Evgeny Kostyuchenko[2], Ekaterina Kosenko[3]
[1]Tomsk State University of Control Systems and Radioelectronics (TUSUR), Tomsk, Russia, ndi@keva.tusur.ru
[2]Tomsk State University of Control Systems and Radioelectronics (TUSUR), Tomsk, Russia, key@keva.tusur.ru
[3]Tomsk State University of Control Systems and Radioelectronics (TUSUR), Tomsk, Russia, key@keva.tusur.ru

ABSTRACT

The work considers a software package for automating speech quality assessment. The complex is used for speech rehabilitation of patients after surgical treatment of oncological diseases of the organs of the speech-forming tract. The structure of the complex and the role of its individual modules are presented. The complex has been introduced into the activities of the Oncology Research Institute of Tomsk Scientific Research Center.

Key words : Speech quality, Speech rehabilitation, Cancer of the mouth and oropharynx

1. INTRODUCTION

Currently, more than 100,000 cases of cancers of the organs of the speech-forming apparatus have been identified [1], and this number is increasing by 25,000 every year [2,3]. The most common treatment option involves surgery, after which it is necessary to undergo speech rehabilitation. The obvious shortcomings of the method of assessing the quality of speech restoration, based on GOST R 58040-95 [4], led to the need to develop algorithms for automatic evaluation and their implementation in the framework of automated systems for assessing speech intelligibility. At the moment, there are a number of works on automation of speech intelligibility assessment [5-7], but all of them appeared later than the beginning of development and the appearance of the main algorithms of the considered software complex. As part of the research on speech restoration using technical methods, such algorithms were developed and implemented. They were included in the developed software package described in [8]. After preliminary testing, it was decided to finalize the software package.

To implement the software package for evaluating the quality of speech in the speech rehabilitation process of patients as part of the cancer treatment at Tomsk Cancer Research Institute, it was necessary to achieve the following objectives:

1. Improvement of the database structure to store several variants of the syllabic intelligibility evaluation and signal segmentation, the ability to specify the type of session (not only syllabic or phrasal intelligibility, but also the mark of the moment of recording the session: before or after the operation);
2. Adding previously developed lists of syllables for recording to the databases.
3. Development and implementation of estimation algorithms based on the previously described approaches [9,10] and segmentation algorithm to evaluate individual phonemes in syllables;
4. Improvement of the syllable recording module: development of more accurate algorithm for detecting voice activity, adding the ability to configure its parameters, and the ability to correct the process of syllables recording;
5. Adding the syllables pronunciation quality evaluation in real time during the recording session.

2. DATABASE

In this article database that is presented in [11] is used. Due to the fact that it is necessary to store several estimates for one syllable entry, 5 fields were added to the Slog table to store estimates (type double), as well as a field to store syllable segmentation (type varchar). Segmentation is a set of values that indicate the boundary between phonemes, separated by a tab. The changes in the Slog table are presented in Figure 1.

Column Name	Datatype
id_zap	INT(11)
id_seans	INT(11)
id_slog	INT(11)
path	VARCHAR(255)
listen	VARCHAR(255)
correct1	DOUBLE(255,4)
correct2	DOUBLE(255,4)
correct3	DOUBLE(14,4)
correct4	DOUBLE(14,4)
correct5	DOUBLE(14,4)
segmentation	VARCHAR(255)

Figure 1: Changes of Slog table structure in the database

Also, a type field (varchar) has been added to the Seans table, where information about all sessions is stored. This allows to add comments or additional information about the

Support by Russian Scientific Foundation, the project "Restoration of speech function using technical methods and mathematical modeling in patients with cancer of the oral cavity and oropharynx after surgical treatment", No. 1615-00038.

session. This field is expected to contain data about the recording period of the session, before or after the operation.

To segment syllables, the position_trouble field (int) has been added to the Hostslog_tr table, which reflects where the problem phoneme is located within the given syllable.

In addition to changes to the database structure, syllable sets that can be used to conduct and evaluate sessions have been added. Currently, there are 5 sets in the database, according to three of which it is possible to evaluate the session. In table 1 the description of syllabic assessment sets is presented.

Table 1: Description of syllable sets in the database

The name of the set	Quantity	Description
GOST 1	250	The first 250 syllables of GOST R 58040-95
Count	10	Numbers from zero to nine
OncoTongue	90	A set of syllables with problem phonemes [k][s][t] and their soft variants: 5 syllables for each possible phoneme arrangement in a syllable (at the beginning, in the middle, at the end)
OncoTongueNK	60	A set of syllables with problem phonemes [k][s][t] and their soft variants: 5 syllables for each arrangement of the phoneme at the beginning or at the end of the syllable
OncoTongueN	30	A set of syllables with problem phonemes [k][s][t] and their soft variants: 5 syllables with the phoneme location at the beginning of the syllable

The OncoTongueNK and OncoTongueN syllable sets were created due to the limitations of the segmentation algorithm (it works only if the phonemes are located at the beginning of a syllable with any number of phonemes in a syllable and if the phoneme is located at the end of a syllable with three phonemes in a syllable). For each of the syllable sets, transcriptions and numbers of the problematic phonemes in them were also added.

3. PARAMETERS AND SETTING OF THE VOICE ACTIVITY DETECTION ALGORITHM

При проведении анализа речевого сигнала при решении различных задач, в частности, выделения голосовой активности, необходимо выделить параметры для дальнейшей обработки [12].

The voice activity detection algorithm used in the program works by calculating two parameters of the signal recorded in the buffer: the spectral flatness measure SMF (1) and energy E.

$$SMF = 10 \times \log_{10} \frac{Gm}{Am}, \quad (1)$$

where *Gm* is the geometric mean of the speech spectrum, *Am* is the arithmetic average of the speech spectrum.

The algorithm makes a decision about the presence or absence of speech in the signal buffer as follows: the parameters for the signal buffer are calculated and compared with the minimum values. If the difference between the received values and the minimum is greater than the threshold values, then a decision is made about the presence of voice in this segment.

In order to be able to adjust the algorithm parameters for various recording devices, a new form was added, which allows record sound without voice ("silence") and voice recording. Based on the obtained parameters, further recording can be carried out. The form is shown in Figure 2.

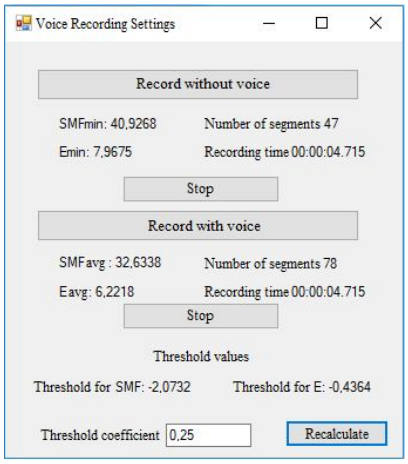

Figure 2: A form designed to configure the parameters of the algorithm for detecting voice activity

The ability to adjust thresholds depending on the microphone sensitivity and patient speech was also added. To do this, a multiplier field was added, which allows the editing the threshold. Currently, the optimal multiplier for sensitive microphones is 0.5, for any other type 0.25. The optimal time for each type of recording is 4-8 seconds.

Setting voice recording options are available both from the main application window (the window with the list of patients) and from the window in which the session is recorded directly.

4. SYLLABLE WRITING MODULE

It is assumed that the recording of the session will take place in an automated mode, namely, a speech therapist (or patient) will only participate in the recording settings, and processing the recordings, calculating the score and switching to the next syllable will be automatic. The syllable form is shown in Figure 3. In this version of the software package, you can configure the following parameters:

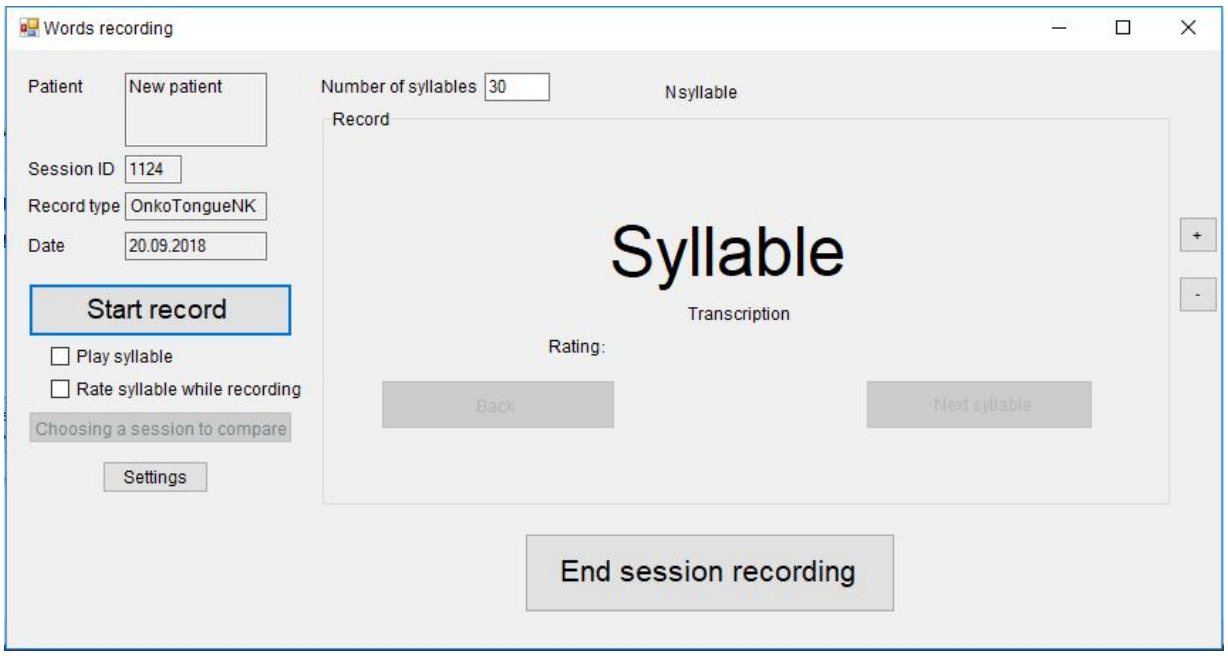

Figure 3: Form for syllables recording

1) The ability to play a syllable. It is possible to play the recorded reference pronunciation of the syllable through headphones or speakers for the sets of syllables OncoTongue, OncoTongueNK, OncoTongueN. This feature should help the patient understand how to pronounce the syllable displayed on the screen.

2) Display the syllable and its transcription on the screen. In the previous version of the program, both the syllable and its transcription were displayed with the same size (approximately 25 size). However, after the first test run of the program at the Oncology Research Institute, it was decided to significantly increase the font for the syllable and decrease for transcription, since the patient did not understand the correspondence between the syllable and transcription. It was also possible to increase or decrease the font size for displaying the syllable.

As part of the software package [8], if the voice activity detection algorithm did not work correctly, returning to the previous syllable was carried out by pressing Ctrl + Z, the "Back" button was added in the new version of the program, which performs the same function. Replacing a keyboard shortcut with a button is justified by the fact that the purpose of a button with a similar inscription is intuitively clear to the user, which simplifies working with the program. The “Next syllable” button remains for recording syllables of patients whose speech has deteriorated after surgery to a state in which it is impossible to correctly determine the beginning and end of speech. In this case, the sensitivity parameters of the algorithm will be changed to the minimum, sound will be recorded continuously, and switching between syllables will be carried out by a speech therapist.

To assess the variability of speech before the operation, it is necessary to record two sessions, which will be used as a reference in the future. However, to familiarize the patient with the software complex and facilitate further work with it, it is proposed to record three or more sessions, and choose only two of them as a reference.

When recording sessions after an operation, it is possible to evaluate the session directly during recording. To do this, select "Evaluate syllables when recording", after which the button "select a session for comparison" will be available. If you select to evaluate but do not select reference sessions, the program will issue a warning and will not start a recording session until the conflict is resolved.

The session selection for comparison is carried out in the form shown in Figure 4. It is possible to select one or two sessions for comparison. The ability to select only one session for the treatment is added due to the fact that there are patients who are currently being treated at The Oncology research Institute, but before the operation they managed to record only one session. It is possible to select only sessions of the same type (with the same set of syllables) as the session being recorded.

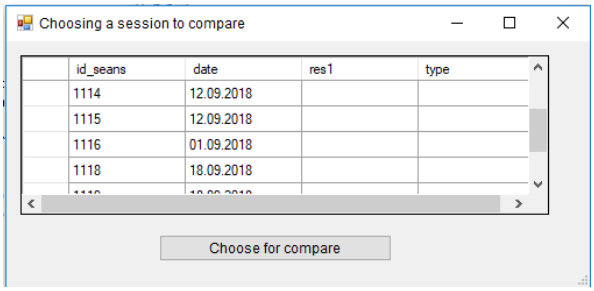

Figure 4: Form for choosing a session to compare

5. SYLLABLE EVALUATION MODULE

In this complex there are three different ways of assessing the syllable intelligibility. They are calculated as the ratio of the average correlation coefficient between the pairs of reference and estimated pronunciation to the correlation coefficient between the reference pronunciations. The situation with two reference sessions is described here and below, but all estimates are also available for 1 session. Thus, the resulting value is enclosed in the range from 0 to 1, and the closer it is to 1, the closer to the standard the phoneme is pronounced. In the first and third methods, the Resample function from the Matlab mathematical package is used to calculate the correlation coefficient, in the second method, the algorithm of dynamic transformation of the timeline is used. The difference between the first method and the third one is that in the first method not the entire syllable is evaluated, but only the problematic phoneme. Therefore, before the direct assessment, it is necessary to segment the syllable and highlight the problematic phoneme. In view of the fact that the segmentation algorithm works correctly only for certain phoneme locations in a syllable, the first method of evaluation is available only for certain sets of syllables.

Syllable Assessment is possible not only in the process of recording a session, but also for an existing session. The form for viewing the session and its evaluation is presented in Fig. 5. In order to carry out segmentation, it is necessary to click the “Perform segmentation” button, if the segmentation function is not available for this type of session, a message will be displayed about this. The segmentation results are stored in the Segmentation column and after clicking the “Save” button are written to the database. Next, you need to select the evaluation settings, namely the type of evaluation and reference sessions. Reference sessions are selected similarly to the syllable entry form. After selecting the reference sessions, the button “Rate Session” will become available and when it is pressed, a sequential evaluation of all syllables in this session will take place. The result will be presented in the correct1 column and after clicking the "Save" button will be saved in the database. After the conclusion of all ratings, the average rating of the session is calculated, and is also entered into the database.

5. CONCLUSION

The paper describes the problem points identified in the process of testing the software complex of speech intelligibility assessment and formulated the objectives for their correction. During the period of work with each of the objectives, the necessary functions were implemented, as well as the currently existing limitations on working with the software complex were described. The modified software package is being implemented at Tomsk Cancer Research Institute, test records were conducted, and the first patient records were made. The result of the analysis of the records were recommendations for further refinement of the software package:

- add fields to a table with patient data to store detailed medical information;
- add the ability to record separated syllables for cases of late detection of problems with the recording session (for example, during additional listening within the syllable detected extraneous noise that interferes with its perception);
- add the ability to display scores obtained in different ways in multiple columns with syllable scores.

Further modernization of the software package involves the use of speech recognition and analysis systems [13-19] to assess syllabic, verbal and phrasal intelligibility. To ensure the security of stored information, it is planned to develop a cryptographic protection module that would not interfere with obtaining real-time ratings [20].

REFERENCES

1. R. Jothikumar, G. Siva Shanmugam, M. Nagarajan, S. Premkumar, A. Asokan **Analyzes of mouth cancer using max-min composition in soft computing** *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 3, pp. 825-830, 2019. https://doi.org/10.30534/ijatcse/2019/76832019
2. A.D. Kaprin, V.V. Starinskiy, G.V. Petrova **Malignant neoplasms in Russia in 2016 (morbidity and mortality),** 2018, 250 p.
3. A.D. Kaprin, V.V. Starinskiy, G.V. Petrova **The status of cancer care for the population of Russia in 2016**, 2017, 236 p.
4. GOST R 50840 **Voice transmission over communication paths. Methods for assessing quality, legibility and recognition**, 1996, 234 p.
5. H. Pamula et al. **Parametric Assessment of Esophageal Speech in Post-Laryngectomy Patients** *2018 Joint Conference-Acoustics, IEEE*, pp. 1-5, 2018.
6. S. Kalita, S.R. Mahadeva Prasanna, S. Dandapat **Intelligibility assessment of cleft lip and palate speech using Gaussian posteriograms based on joint spectro-temporal features** *The Journal of the Acoustical Society of America*, vol. 144, no. 4, pp. 2413-2423, 2018. https://doi.org/10.1121/1.5064463
7. Y.I. Sumita et al. **Digitised evaluation of speech in-telligibility using vowels in maxillectomy patients** *Journal of oral rehabilitation*, vol. 45, no. 3, pp. 216-221, 2018. https://doi.org/10.1111/joor.12595
8. E. Kostyuchenko et al. **Software for the objective assessment of the quality of pronunciation of syllables in speech rehabilitation** *All-Russian Scientific Conference on Management Issues in Technical Systems. - Federal State Autonomous Educational Institution of Higher Education St.*

Petersburg State Electrotechnical University LETI named after VI Ulyanov (Lenin), vol. 1, pp. 277-280, 2017.

9. E. Kostyuchenko et al. **Correlation normalization of syllables and comparative evaluation of pronunciation quality in speech rehabilitation** *International Conference on Speech and Computer*, Springer, Cham, pp. 262-271, 2017. https://doi.org/10.1007/978-3-319-66429-3_25
10. D. Novokhrestova **Temporary normalization of syllables by the dynamic transformation algorithm of the time-line in assessing the quality of the pronunciation of syllables in the process of speech rehabilitation** *Proceedings of TUSUR*, vol. 20, no. 4, pp. 142–145, 2017.
11. E. Kostyuchenko, D. Novokhrestova, A. Pyatkov **Formation of a database of patients with speech rehabilita-tion after combined treatment of oncological diseases of the organs of the speech-forming tract** *Electronic means and control systems*, Tomsk, vol. 2, pp. 245-247, 2017.
12. M.A. Mazumder, R.A. Salam **Feature extraction techniques for speech processing: A review** *International Journal of Advanced Trends in Computer Science and Engineering*, *v*ol. 8(1.3 S1), no. 54, pp. 285-292, 2019. https://doi.org/10.30534/ijatcse/2019/5481.32019
13. I.S. Kipyatkova, A.A. Karpov **Variants of Deep Artificial Neural Networks for Speech Recognition Systems**. *Trudy SPIIRAN – SPIIRAS Proceedings*, vol. 6, no. 49, pp. 80-103, 2016. https://doi.org/10.15622/sp.49.5
14. I.S. Kipyatkova, A.A. Karpov **Analytical review of Russian speech recognition systems with a large dictionary** *Trudy SPIIRAN – SPIIRAS Proceedings*, vol. 12, no. 1, pp. 7-20, 2010. https://doi.org/10.15622/sp.12.1
15. K. Ishikawa, J. MacAuslan, S. Boyce **Toward clinical application of landmark-based speech analysis: Landmark expression in normal adult speech** *The Journal of the Acoustical Society of America*, vol. 142, no. 5, pp. EL441-EL447, 2017. https://doi.org/10.1121/1.5009687
16. M.S.Hossain, G. Muhammad **Emotion recognition using secure edge and cloud computing** *Information Sciences*, vol. 504, pp. 589-601, 2019. https://doi.org/10.1016/j.ins.2019.07.040
17. Heysem Kaya, Alexey A. Karpov, and Albert Ali Salah. **Fisher vectors with cascaded normalization for paralinguistic analysis**. *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.
18. S.K. Rososhek, R.V. Meshcheryakov, A.A. Shelupanov, M.A. Sonkin **Cryptographic protocols in systems with limited resources** *Computational Technologies*, vol. 12, no. S1, pp. 51-61, 2007.
19. I. Rakhmanenko, A. Shelupanov, E**.** Kostyuchenko **Fusion of biLSTM and GMM-UBM systems for audio spoofing detection** *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 4, pp. 1741-1746, 2019. https://doi.org/10.30534/ijatcse/2019/103842019
20. E. Kostuchenko, D. Novokhrestova, M. Tirskaya, A. Shelupanov, M. Nemirovich-Danchenko, E. Choynzonov, L. Balatskaya **The evaluation process automation of phrase and word intelligibility using speech recognition systems** *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11658 LNAI, pp 237-246, 2019. https://doi.org/10.1007/978-3-030-26061-3_25