

## **Acoustic Simulator for the Hearing Impaired People**

**Farkhadov M. P. doctor of technical sciences**

Moscow, Russian

Russia Institution of the Russian Academy of Sciences  
Institute of Management Issues named after V. A. Trapeznikov  
Russian Academy of Sciences

**Kudubayeva S. A. candidate of technical sciences, docent**

Kostanay, Kazakhstan

AkhmetBaytursynovKostanay State University,

**ErmagambetovaG. N. m.sc.**

Kostanay, Kazakhstan

AkhmetBaytursynovKostanay State University

### **Annotation**

*In the given article the purpose and working principle of acoustic simulator, software architecture of the system, its characteristics are considered.*

**Key words:**hearing impaired people, acoustic simulator, speech recognition, interface, software architecture of the system.

### **1. Introduction**

The most effective way of teaching hearing impaired people the correct speech is lessons with experienced audiologists. Nevertheless, in order to help people in violation of articulation as well as hearing impaired people there are computer simulators, the appearance of which has become noticeable in recent years. It is aided by the presence in such simulators of the visual feedback function which is of utmost importance for the hearing impaired and deaf people. The first computer speech simulators were designed by the IBM at the end of 1980s. These simulators, which were called SpeechViewers [1], demonstrated their usefulness in the USA and other English-speaking countries and contributed to the emergence of this product's versions for other languages including Russian: "Visible speech" program [2]was a Russian version of SpeechViewer. There are other such programs, for example Indiana Speech Training Aid (ISTRA) [3], Speech Illumina Mentor (SIM) [4], Speech Training, Assessment, and Remediation system (STAR) [5], ArticulationTUtoR (ARTUR) [6]. Simulators' working principle is in the spectral transformation of the audio input, identification of its characteristics, its comparison with the standard, its return and presentation of the result. In some simulators to analyze an initial signal automatic speech recognition systems are used. All these tools presented independent software that had to be purchased and installed on the user's computer.

Placement of such programs in the Internet and providing for the online access to them would let to drastically increase a number of people, who have the possibility to train the correct pronunciation of sounds, syllables, words and phrases by themselves using home computers. Especially for this purpose works on acoustic simulator were begun.

### **2. Acoustic simulator's interface design, its purpose and working principle**

In Kazakhstan, according to the information provided by the Ministry of Labour and Social Welfare of the Republic of Kazakhstan, there were about 21 000 hearing impaired people in the beginning of 2013. Their inclusion in the labor and social life is closely connected with their teaching oral verbal speech. As a rule, speech apparatus of such people is not disrupted but they cannot use it on their own without special training.

The purpose of the designed computer simulator was not the audiologist's replacement but providing users with additional opportunities for self-study, which are especially important while teaching hearing impaired children or children in violation of articulation, because parents can join the teaching process and children will get the possibility to train articulation at home, not only in the teacher's room.

As a result the "Acoustic simulator" prototype was designed and it let working via the local network. Moreover, new Internet service that of acoustic simulator is being considered to teach online the correct pronunciation of sounds of the Russian and Kazakh languages.

Simulator's working principle lies in the following: encoded acoustic stream is sent from the user's computer to the media server, where it is decoded into the required format. Then the sound is sent to the decoding server, which returns the degree of correctness of the spoken sound. Media-server forms an image of the pulse-code modulation of the spoken sound over the standard version and returns it to the client together with the information about the degree of its correspondence.

Placement of such programs in the Internet and providing for the online access to them increases drastically a number of people, who get the possibility to train the pronunciation of sounds using home computers. Besides the hearing impaired people this simulator can be used by hearing people as well as by people, for whom the Russian and Kazakh languages are not native.

To solve the problem, it was decided to use computer speech recognition as an analyzer of audio signal coming from the user. The recognizer's core and corresponding processing and control programs are located in the central part of the processing complex hosted on a server. Users get necessary images, instructions and hints on their screen. Special reference information provides users with the general notion about sounds and with the detailed description of the way how to pronounce every single sound. The basis for the software architecture of the "Acoustic simulator" application is made by the "client-web-server-application server" concept.

According to the researches made the hearing impaired people get all the most important information watching lips' movements. In this connection it was decided to place on the screen two windows, in which the lips' movement (full face and profile view) is demonstrated while pronouncing the chosen sound. There is also a window, where you can watch the animated image of the speech apparatus for the chosen sound. Textual description of the vocal tract's working principle is given, the possibility to demonstrate the sound by pushing the corresponding button is provided. The user is also given the opportunity to control his lips' movements while pronouncing the chosen sound, if his computer is equipped with web-camera. To provide the feedback sound visualization is possible: an image of the pulse-code modulation of the standard version's sound and the spoken sound as well as the mark for the pronunciation accuracy is displayed.

After recording and sound analysis the results consisting of the following points will be displayed:

- Mark for the correspondence of the spoken sound in its relation to the standard version's sound on a 100-point scale;
- One of five images corresponding to the mark on a 5-point scale;
- Images of the pulse-code modulation of the spoken sound (red color) and standard version's sound (blue color).

This simulator to train and correct sounds pronunciation with online access via the Internet is a new and original service in the Internet of a new generation. Development of the simulator's functionality goes in the following directions: adding the function of diphthongs, syllables, words and phrases training, working out the methods of the pronunciation adjustment, maintaining an individual student's diary and creating an individual training course, which adapts to the user in the learning process, creating services for those teachers who use this simulator.

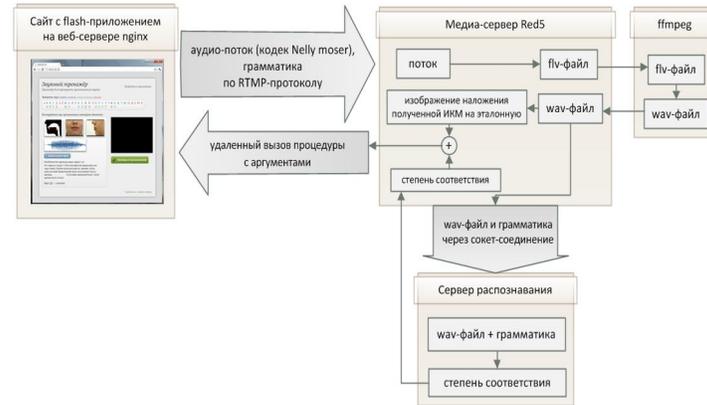
### **3. Software architecture of the system**

On picture 1 software architecture of the application, in the basis of which the "client-web-server-application server" concept lies, is presented.

The system contains the following components:

- Application designed on the Adobe Flash platform and embedded in a web-page is the client.
- Program framework of Adobe Flex 4 client's application.
- Nginx web-server.

- Red5 media-server is the application server.
- Nuance Recognizer 9 is the speech recognizer.
- In the course of application's work server also uses software for the ffmpeg media-information's encoding and decoding.
- 



**Picture 1 – Architecture of the system**

Adobe Flex 4 framework, nginx and Red5 servers, ffmpeg are software with an open source code, which are, respectively, distributed for free.

#### **4. The technical solution and characteristics**

Technical solution's description is best illustrated in the form of system elements' interaction chain (see picture 1):

1. When the user enters the site (for example, [speech-trainer.asmon.ru](http://speech-trainer.asmon.ru)), browser sends http-request to the nginx web-server to obtain the index.html page.
2. On the page obtained there is the code loading the swf-file of the flash-application from the web-server onto the user's computer.
3. After the loading of the swf-file of the flash-application initialization of the application by passing it to the AVM (ActionScript Virtual Machine) for execution takes place.
4. The user chooses the necessary letter on the alphabet panel. Alphabet panel is equipped with the color marking of different types of sounds, which are correlated with the corresponding letters.
5. Sound information corresponding to the chosen letter is loaded from the XML-file embedded in the application:
  - a) One swf-animation presenting the vocal tract's work,
  - b) Two videos in the flash video format (flv-files) presenting sound pronunciation performed by a man in a full face and profile view,
  - c) An image of the pulse-code modulation of the sound,
  - d) Textual information about the way of forming a sound (or sounds, if the letter consists of two sounds, for example "Yo", "Ya") and sound's short characteristic.
6. When pushing the "Demonstration" button animation is launched and videos are played once.
7. When pushing the "Check the pronunciation" button over the application's interface a panel (panel's image) appear. This panel guides the user in the process of recording and reviewing the results of the pronunciation of the chosen letter:
  - a. First of all you are offered to pronounce a sound (or sounds) corresponding to the chosen letter. Application is connected with the remote server with the help of a socket,
  - b. In the very beginning of the pronunciation audio-stream transmitted to the media-server and encoded with the help of the NellyMoser's proprietary audio-codec opens. The use of this codec helps to minimize the traffic between the client and the media server,

- c. After the completion of the sound pronunciation the audio stream is closed and the cycle of the audio stream's processing begins:
- audio-stream is saved in the flv-file hosted on the media-server;
  - with the help of the ffmpegsoftware the extraction of the audio-track from the flv-file, its decoding and saving in the wav-file take place;
  - the way to the formed wav-file as well as the sound's name, which defines the active grammar, are transmitted through the socket-connection to the recognition server's entrance. This server is based on the Nuance Recognizer 9 core. Acoustic file is checked for the correspondence to the active grammar;
  - at the output the recognition server generates a numerical value indicating the degree of correspondence of the sound transmitted to the mentioned grammar in the range from 0 to 100. This meaning is returned by the recognition server to the media-server;
  - At the same time media-server creates an image of the pulse-code modulation of the sound stream. A created image is superimposed on the pulse-code modulation of the standard version. An obtained image is saved;
  - An obtained meaning (the degree of correspondence) and the created image of the produced pulse-code modulation having been superimposed on the pulse-code modulation of the standard version is returned by media-server to the client by invoking the corresponding remote (in relation to the media server) procedure.
- a. The client, having received such data defines the degree of correspondence, a visual image in the form of "smile" pictures (5 images for the degrees ranges of 0-20, 21-40, 41-60, 61-80, 81-100 correspondingly are used) as well as an image of the obtained pulse-code modulation.
- Further the user may close the results panel, try again to pronounce the sound corresponding to the chosen letter or choose another letter.

### **Conclusion**

"Acoustic simulator" web-application is designed and launched within the laboratory's computer network. Application is ready to be placed in the Internet. In order to implement this hosting on the cloud platform with the possibility to install Red Hat Linux OS is necessary. The choice of this OS is stipulated by the presence of the Nuance Recognizer version just for this Linux distribution kit. It is also necessary to purchase the commercial license to make the work of Nuance Recognizer system of automatic speech recognition possible.

Further development of the simulator's functionality (adding the function of diphthongs, syllables and words training, working out the methods of the pronunciation adjustment, creating services for those teachers who use this simulator) seems to be extremely essential.

### **References**

- Adams, F.R., Crepy, H., Jameson, D. and Thatcher, J., 1989, IBM Products for Persons with Disabilities. Global Telecommunications Conference and Exhibition. 'Communications Technology for the 1990s and Beyond', GLOBECOM '89, IEEE, 2, pp. 980 – 984.
- <http://logoped.far.ru/p54aa1.html>
- Watson, C., Reed, D., Kewley-Port, D. and Maki, D., 1989, The Indiana speech training aid (ISTRA). Comparisons between human and computer-based evaluation of speech quality. *Journal of Speech and Hearing Research*, 32, pp. 245 – 251.
- Soleymani, A.J.A., Mccutcheon, M.J. and Southwood, M.H., 1997, Design of speech mentor (SIM) for teaching speech to the hearing impaired. In *Proceedings of the Sixteenth Southern Biomedical Engineering Conference*, 4 – 6 April 1997, Biloxi, MS, pp. 425 – 428.
- Bunnell, H.T., Yarrington, D.M. and Polikoff, J.B., 2000, Star: articulation training for young children. In *Proceedings of the 6<sup>th</sup> International Conference on Spoken Language Processing*, 4, 16 – 20 October 2000, Beijing, China, pp. 85 – 88.
- Engwall, O., Wik, P., Beskow, J. and Granstroy M, B., 2004, Design strategies for a virtual language tutor. In *Proceedings of the 8<sup>th</sup> International Conference on Spoken Language Processing*, 4 – 8 October 2005, Jeju Island, Korea, pp. 1693 – 1696.