

#99

The electroencephalogram based classification of internally pronounced phonemes

Yuliya Gavrilenko, Saada Daniel, Eugene Ilyushin, Alexander Vartanov, Andrey Shevchenko
Lomonosov Moscow State University, Moscow, Russia



Abstract

The internal speech recognition is a promising technology, which could find its use in brain-computer interfaces development and greatly help those who suffer from neurodegenerative diseases. The research in this area is in its early stages and is associated with practical value, which makes it relevant. It is known that internal pronunciation can be restored according to elec-troencephalogram data because it allows one to register specific activity associated with this process. The purpose of this work is to build and implement an algorithm for extracting features and classifying Russian phonemes according to an electroencephalogram recorded during the internal pronunciation of the phonemes. This kind of research is actively conducted abroad; however, at the moment, there is no information in open sources about such works for the Russian language phonemes. In the course of the work, an algorithm for extracting features and classifying the internal pronunciation of Russian phonemes was built and tested, the accuracy of which showed results comparable with other studies.

Experiment

An experiment was conducted in the psychophysiology laboratory of the Faculty of Psychology of the Lomonosov Moscow State University to obtain data from internal pronunciation. The study involved 12 subjects: 3 females and nine males aged 20 to 25 years (average age - 22 years). All subjects did not have any traumatic brain injury or mental illness and were right-handed. Two types of experiment were conducted:

1. An experiment with the visual stimuli presentation;
2. An experiment with the presentation of sound stimuli.

In each of the experiments performed, the cognitive task consisted of prolonged internal pronunciation of a certain phoneme without imaginary articulation. For the experiment, we selected seven phonemes of the Russian language from different subgroups of the Plotkin classification [11]: /a/, /b/, /f/, /g/, /m/, /r/, /u/.

To present a visual stimulus, the Presentation program (version 18.0 of Neurobehavioral Systems, Inc.) was used, which laid out the necessary algorithms for giving stimulus material. For each phoneme, using a photo editor, an image of the PNG format was made, which contained the recording of this phoneme. The received files were loaded into the “Presentation” software complex and presented in random order.

For the presentation of the sound stimulus, headphones were worn, in which subjects listened to the phonemes required for pronunciation pre-recorded on the recorder. Registration of electrical activity of the brain was performed using a 19-channel Neuro-KM electroencephalograph ("Statokin" company, Russia), with a sampling rate of 1000 Hz. The electrodes were located according to the 10-20% international system, a detailed description of which can be found in . BrainSys (BrainWin), version 2.0 was used for recording and viewing EEG. The scheme of the experiment corresponds to the standard schemes described in similar studies:

1. For 700 ms, the subject is shown a visual or audible stimulus, which is an in-struction for internal speaking;
2. After the stimulus is presented, a fixation cross is displayed on the screen on which the subject should focus his gaze;
3. During 1500 ms the screen displays a fixation cross, and the subject performs an internal pronunciation of the phoneme;
4. The fixation cross disappears from the screen and a period of rest begins, during which the subject takes no action and prepares for the next stimulus;
5. After 1600 ms, the rest is completed, and a new stimulus is presented to the subject

Results & Discussion

After applying the SVM-RFE algorithm to the extracted features, the following ranked list was obtained, in which the significance of the feature decreases in the next order: mean value, median, standard deviation, maximum, sum, minimum. Using the cross-correlation method, it was found that the most optimal is the use of the first three signs from this list.

When compiling an array of samples from recordings of experiments with different people, the classification accuracy tends to be random, which is comparable with the studies of other authors. However, when classifying internal pronunciation for only one subject, the results are significantly improved. Below are the results of the pairwise classification of different phonemes for two different subjects (Table 1, 2).

Table 1. The results of the pairwise classification of phonemes for the subject №1

	/b/	/f/	/g/	/m/	/r/	/u/
/a/	0.71	0.48	0.65	0.75	0.63	0.56
/b/		0.47	0.79	0.77	0.67	0.51
/f/			0.61	0.68	0.68	0.74
/g/				0.77	0.81	0.74
/m/					0.63	0.78
/r/						0.67

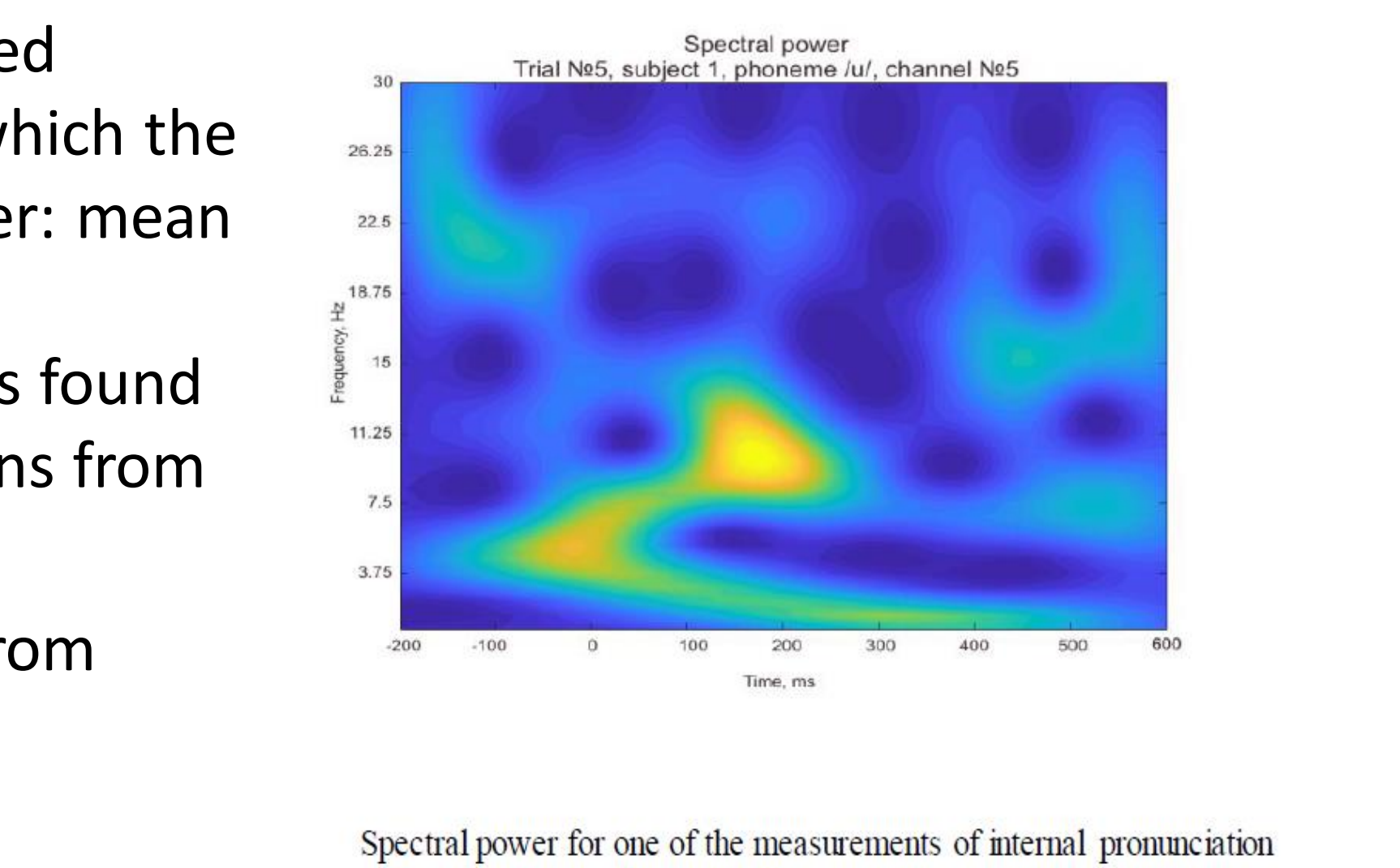


Table 2. The results of the pairwise classification of phonemes for the subject №2

	/b/	/f/	/g/	/m/	/r/	/u/
/a/	0.42	0.53	0.53	0.69	0.60	0.36
/b/		0.57	0.62	0.54	0.62	0.66
/f/			0.47	0.63	0.45	0.57
/g/				0.68	0.64	0.38
/m/					0.74	0.36
/r/						0.54

Conclusions

Due to the control of the data recording during the experiment, the entry does not initially contain artifacts of blinking and muscle movement during the execution of a cognitive task. This method of data cleaning does not allow using the constructed algorithm for real-time classification but provides proper filtering of data from artifacts. At the current stage of development of internal pronunciation studies, such an approach is quite acceptable, and the constructed algorithms can later be adapted for online classification.

As can be seen from the results, the classification accuracy is comparable with the studies of other authors – studies, reviewed earlier, provide accuracy that is alike. It is still not high enough to attempt to create an internal pronunciation based BCI. As we suppose, one the main reasons for the low accuracy is the standard setup of the experiment. Due to the small number of electrodes and installation according to the standard “10–20%” system, only four sensors turned out to be in the area of interest to the brain. We suppose that slightly modifying the setup with considering the experiment features could help increase the recognition accuracy. In future work we are planning to use the 64-channel EEG in order to locate more sensors in the area of interest of the brain. In addition, the use of the fixation cross as an incentive to start internal speaking was a relatively stable source of noise signals. In order to eliminate this drawback, instead of a fixation cross, it is proposed to use a light sound stimulus (for example, a click) and conduct the experiment entirely with closed eyes in the following work.

Even though in the formulation of the experiment, phonemes from various subgroups of the Plotkin classification are used, some pairs are discriminated on average better than others are. The causes of this phenomenon at the moment cannot be determined unequivocally. However, it can be assumed that phonemes from different groups are discriminated better than phonemes from different subgroups of the same group. As an example, the pairs /g/ - /r/, /a/ - /m/ and some others can be given. In the subsequent work, it is planned to increase the stimulus material and to include words in it. The algorithm, that is used in this article, can be applied to other tasks related to BCI construction. On the basis of the researches carried out further works on studying of work of various departments of a brain and their interactions in the course of the decision of intellectual problems by the person are planned.

The research is financially supported by the Russian Science Foundation, Project №20-18-00067

contact: a_v_vartanov@mail.ru