categories. For example, the word traurig (sad) refers to the categories «Affect», «Negative emotion» and «Sadness». We analyzed all the monologues, calculated the number of words for each category and divided them by the total amount of words in each monologue. By this way we got 64 characteristics of 56 monologues. Our task was to investigate the dependency between these 64 features and the results of the subtest «Information», and to find several informative features out of 64 characteristics.

We combined 4 or 5 features together, trained the non-parametric model and applied our method. As a result, the category «Affection» had the largest value of the first-order partial derivative and was estimated as a more informative feature. «Positive emotions» and «Negative emotions» are subcategories of «Affection» and are also relevant according to our algorithm. However, «Anger» and «Optimism» do not have large values of $\hat{f}'_{x_i}$. The category «Cognitive mechanism» is estimated as irrelevant, however, the category «Cause» which is a subcategory of «Cognitive mechanism» is more important.

**Discussion and Future work.** The goal of this work was to apply the method to the corpora. In each combination of the features the category of emotions was determined as the most informative feature. It means that there is a dependency between speaker's general knowledge and the amount of emotional words which he uses in his speech. We could not find any references describing this dependency. Only in LEAS [6] emotional intelligence is measured linguistically, however, the cor-

relation between them was not found. For our algorithm we used a small data set that also influenced the results. Also, these emotional words may be a subcategory of another category which was not analyzed. For example, they may create a group of frequently-used words, or they are formed from abstract words which show the level of intelligence in spoken utterances. This research and the results are preliminary; in our future work we are going to further investigate this phenomenon, to find other linguistic features which reflect verbal intelligence and to collect more data for more precise estimations.

### References

1. Zablotskaya K., Walter S., Minker W. Speech data corpus for verbal intelligence estimation. In Proceedings of International Conference on Language Resources (LREC) 2010.

2. Wechsler D. Handanweisung zum Hamburg-Wechsler-Intelligenztest fuer Erwachsene (HAWIE). Separatdr., Bern; Stuttgart; Wien, Huber. 1982.

3. Nadaraya E. On estimating regression. Theory of Probability and its Applications. 1964. Vol. 10. P. 186–190.

4. Watson G. Smooth regression analysis. Sankhya – The Indian Journal of Statistics. 1964. Vol. 26. P. 359–372.

5. Computergestuetzte quantitative textanalyse / M. Wolf, A. B. Horn, M. R. Mehl et al. // Diagnostica, 54. 2008. Heft 2. P. 85–98.

6. Lane R. D, Schwartz G. E. Levels of emotional awareness: a cognitive-developmental theory and its application to psychopathology. Am J Psychiatry. 1987. P. 113–143.

S. Zablotskiy, T. Müller, W. Minker

## ESTIMATION OF RADIO SIGNAL QUALITY DEGRADATION BY MEANS OF NEURAL NETWORK AND NON-PARAMETRIC REGRESSION MODEL

*In this paper we present an approach which allows us to avoid expansive and time consuming subjective assessments of audio quality degradation caused by different nature distortions while transmitting and receiving of stereo audio signal through the radio channel. This approach is based on the basic version of PEAQ (Perceptual Evaluation of Audio Quality) originally developed mainly for audio codec estimation. The MOV (Model Output Variables) vector of the PEAQ method is mapped to the audio quality degradation scale using two different models: neural networks and non-parametric regression. The results of two independent approaches are compared.*

*Keywords: PEAQ, audio quality degradation, neural network, non-parametric regression.*

The manufacturers of radio receivers and other radio equipment have to estimate the quality of the new product comparing to the existent equipment. Among other things the common listening comprehension of the perceived degradation with respect to the original (reference) audio signal has to be taken into consideration because humans with their own listening comprehension are supposed to be the main end-users of the designed equipment. Since

any high-quality reliable subjective assessments are very expensive and time consuming it is strongly desired to have a tool for automatic perceptual evaluation of the audio quality degradation. This is a fundamental idea behind the PEAQ method, as specified in ITU-R BS.1387 recommendation [1]. According to this recommendation the PEAQ measurement method is applicable to the most types of audio signal processing equipment, both digital

and analog. However, it is expected that many applications will focus on audio codecs. Test results obtained with the help of the PQevalAudio software, compliant with the basic model of PEAQ [2], disagree with the Mean Opinion Score (MOS) of the test persons. The Objective Difference Grade (ODG) of this software after appropriate linear rescaling is always significantly smaller than MOS. This implies that the PQevalAudio estimates the quality of the received (degraded) audio signal to be permanently lower than it seems to be for subjects from a test group. One of the obvious reasons is that the human requirements for audio files (signals) degraded by audio coding (recoding) are significantly higher than the requirements for files degraded by the entire radio transmitting-receiving process.

The output of the software is not only the final ODG value, but also 10-dimensional Model Output Variables (MOV's) combined using a neural network to give the ODG value [2]. Our goal is to create a new model having MOV's as the input variables and to train it on the basis of our test group subjective estimations to be able to use the whole algorithm for automatic perceptual evaluation of transmitted through the radio channel audio quality degradation. We have chosen two different approaches for MOS approximation: neural network [3] and non-parametric regression model [4]. The performance of both algorithms on the test set is compared and the conclusions are presented in the corresponding section.

**Alignment of Reference and Received Audio Signals.** All test audio files are divided into two unequal parts. The smaller one consists of artificially simulated audio signals degraded by the additionally applied noise signals of different types, which sound more or less real comparing to the radio noise. The second part of the test audio material was obtained under conditions maximally close to the real situation – the reference short audio files were captured from various Internet radio stations with acceptable quality which were aired at the same time (with a small time delay) in our local area. Degraded audio files were captured from the corresponding radio stations in different conditions: inside the buildings and outside, in the driving car.

PQevalAudio software assumes two input audio signals (reference and degraded) to be time and gain aligned. Therefore, all the files are aligned even before the subjective assessment stage to be more accurate, when finding the dependency between MOS and MOV's.

*Variable Delay Compensation.* The reference and degraded files transmitted through the radio channel have not only a constant delay between each other, but also a variable delay changing through the entire audio signal. The constant delay is calculated as a maximum value of the cross-correlation function. The variable delay compensation is an increasingly complex process. Both reference and degraded stereo files are first converted into monaural ones and then divided into blocks of chosen length. For each pair of blocks the mean delay is found. This vector is then digitally filtered by the use of the normalized symmetric Hanning window. The result delay vector is then linearly interpolated for each sample and

the final dynamic shift is performed by the cubic interpolation.

Usually the variable delay seems to have close to linear dependency from the time. However, sometimes due to the strong sudden noise or other distortions there are areas of the variable delay graph, where the curve seems to be almost randomly changing. In this case it is recommended to approximate linearly such «unreliable» areas using the left «trustful» data (see Fig. 1).
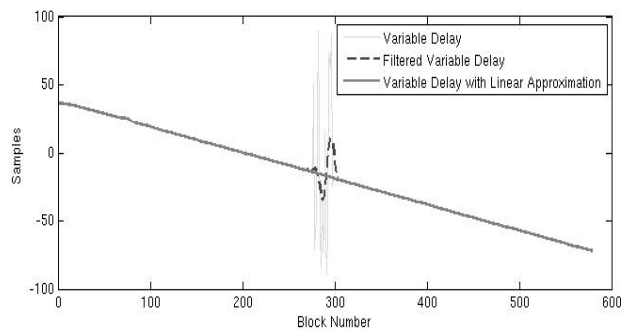


Fig. 1. Variable delay between reference and degraded files with and without approximation

In some specific cases the graph of the variable delay seems to change randomly. It is statistically estimated that in this case it is preferable to ignore the variable delay vector. The recognition of such case is automatically performed according to the acceptance threshold 0.1, i. e. the number of «reliable» data should not be less than 10 % of all block delays data. The smaller threshold would lead to the increasing of the average difference between files.

*Gain Alignment.* As a first step the reference and degraded stereo audio signals are converted into the monaural files. This should take into account, that the most of the receivers switch the audio to mono in the case of poor radio signal. Therefore, processing on the stereo signal may lead to a wrong gain calculation because of the reference and degraded files difference, caused even only by the switching to monaural signal, which can be quite pure itself.

The gain itself is calculated as the absolute value average of all samples of the monaural files. The sample values of the file yielding a higher gain are linearly reduced according to the gain difference. Then, the proportional normalization is applied for both files.

**Subjective Assessments.** Aligned files have to be estimated by the candidates from our test group. They should listen at first the reference and corresponding degraded audio files and then estimate the degradation of the second file with respect to the reference one. In our case each candidate has estimated 58 pairs of files, giving for every pair one value from the interval [0, 10] according to Table.

All pairs of files and appropriate MOS values are divided into two sets: training set used for model creation and the test set (10 pairs from 58). The data from the test set are not used for modeling to be honest in evaluating of approximation results.

**Subjective estimation scale**

| Grade | Subjective impression |
|---|---|
| 0 | Program content is not/almost not understandable |
| 2 | Program content is in common understandable, irritative noise |
| 4 | Program content is mostly good understandable, especially when concentrating |
| 6 | Noise is (always) audible, also at the usual loudness |
| 8 | Reception distortion is audible when concentrating or at the high loudness |
| 10 | Reception distortion is inaudible even concentrating or at the high loudness |

**Neural Network Training.** The modified approach for neural network training [5] by means of the genetic algorithm [6] is used. Genetic algorithms are global optimization stochastic algorithms which do not need the knowledge about the function and its derivatives. Only the values of the optimized function in the generated points are required for optimization [7; 8].

In our approach the genetic algorithm is applied two times within each iteration. Firstly it is used to optimize the weights of all neural nets in the current generation. The optimum in our case is the minimal value of the mean squared error (MSE) between the neural network output and the corresponding MOS values from a training set. Secondly the genetic algorithm is used to determine the new generation of the neural nets. In order to apply the optimization algorithm to a network structure its binary representation is used. The training is iterated until the stop conditions are satisfied or the break is performed manually.

**Non-parametric regression model.** Non-parametric kernel regression of a static object allows us to build a model of this object without any knowledge about the structure of the dependency between input (MOV's) and output (MOS) variables. This model can be presented in the form of a next function:

$$\hat{y}(\vec{x}) = \frac{\sum_{i=1}^{s} y[i] \prod_{p=1}^{n} \Phi\left(\frac{x_p - x_p[i]}{c_p}\right)}{\sum_{i=1}^{s} \prod_{p=1}^{n} \Phi\left(\frac{x_p - x_p[i]}{c_p}\right)}, \qquad (1)$$

where $\Phi(\cdot)$ is a kernel function; $c_p$ — smoothness parameters, $p = \overline{1,n}$; $\vec{x}[i], y[i]$ — input and output variables from the training set, $i = \overline{1,s}$.

The kernel function should satisfy certain requirements [9] and in our case is chosen to be the following function:

$$\Phi(z) = \begin{cases} (3/4)(1-z^2), & if\ |z| \le 1; \\ 0, & if\ |z| > 1. \end{cases} \qquad (2)$$

The only unknown parameters of the algorithm are smoothness parameters $c_p$ which should be trained on the base of the available data, minimizing Mean Square Error (MSE):

$$MSE(c_p) = \left( \frac{1}{s} \sum_{i=1}^{s} \left( \frac{y[i] - \sum_{\substack{j=1, \\ j \ne i}}^{s} y[j] \prod_{p=1}^{n} \Phi\left(\frac{x_p[i] - x_p[j]}{c_p}\right)}{\sum_{\substack{j=1, \\ j \ne i}}^{s} \prod_{p=1}^{n} \Phi\left(\frac{x_p[i] - x_p[j]}{c_p}\right)} \right)^2 \right)^{\frac{1}{2}} . \quad (3)$$

To minimize this error one could use any optimization methods, especially stochastic methods aiming to find a global minimum, e. g. genetic algorithm. It is also possible to chose one single value of $c$ and calculate all the smoothness parameters in the form of $c_p = c\,(\max_i x_p[i] - \min_i x_p[i])$. The parameter $c$ should be decreased from a fixed value (for instance, $c = 0,5$) with a fixed small step until the value of the MSE starts to increase. To get a better result the found values of $c_p$ should be adapted. We did a post-adaptation in our work manually; however, one could prefer a certain optimization method.

**Experimental results.** As it was previously mentioned, linearly mapped ODG value to the interval [0, 10] tends to be significantly smaller than MOS of the candidates from a test group (the MSE = 5,428).

The MSE of the difference between the output of the found neural network and MOS values calculated on the training data set is equal to MSE = 1,040 and calculated on the test data set is even smaller MSE = 0,971. The result approximation is at least much closer to the MOS, than linearly rescaled ODG software output.

Non-parametric kernel regression model on the same training data has achieved a little bit worse results MSE = 1,091 comparing to neural network. However, on the not-overlapping test data set the non-parametric model performed even better (MSE = 0,910) than our neural network. Moreover, it was found out that 6 out of the 10 sub-optimal smoothness parameters (according to 10-dimensional MOV's) are much larger than the scale of the data in the corresponding dimensions. This means that these 6 input variables do not play any role in the non-parametric model and therefore could be omitted. We created a new non-parametric model on the base of the reduced data size (6 variables were ignored) with the same left smoothness parameters values (re-training of the parameters did not bring any improvement). The result model achieved MSE = 1,092 on the training set and MSE = 0,906 on the test set that is the smallest achieved error on this test set.

The results of both algorithms, rescaled ODG value and the perceptual estimations of the candidates (MOS) for 10 values from a test set are presented in Fig. 2. The closer the approximation to MOS is, the better the model is. Non-parametric regression presented on the graph uses only 4-dimensional input data.

Conclusion and Future Directions. Either the neural network or non-parametric regression model in combination with the PQevalAudio software can be used for the perceptual evaluation of the audio quality degradation, transmitted through the radio channel. The both approaches are obviously not as precise as the

subjective assessments. However, they could be used for the mean opinion score approximation.
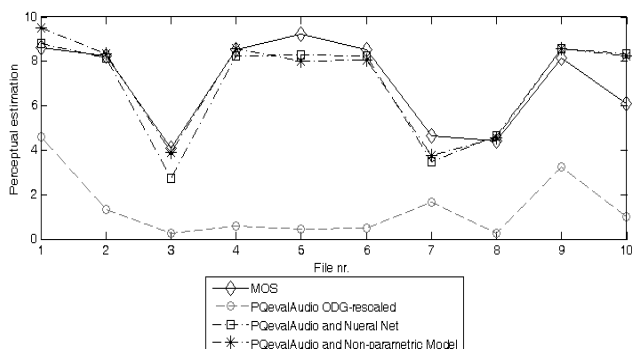


Fig. 2. The algorithms comparison

The non-parametric model has achieved slightly better results on the test set. The main advantage of this model is that we were able to exclude the insignificant features out of the model due to the huge values of the smoothness parameters. The input parameters left in the model are: «average block distortion», «distortion loudness», «noise-to-mask ratio» and «windowed modulation difference». The input parameters which were excluded from a model: «average modulation differences», «bandwidths of the reference and test signal», «harmonic structure of the error» and «relatively disturbed frames». The accuracy of the MOS approximation can be improved by the use of described algorithms trained on the considerably greater amount of data and estimated by a higher number of people.

## References

1. ITU-R Recommendation BS.1387, Method for Objective Measurements of Perceived Audio Quality, Dec. 1998.

2. Kabal P. An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality // Department of Electrical & Computer Engineering, McGill University, Canada, 2002.

3. Anderson J. A. An Introduction to Neural Network, The MIT Press, 1995.

4. Nadaraya E. On estimating regression, Theory of Probability and its Applications. 1964. Vol. 10. P. 186–190.

5. Hagan M. T. Neural Network Design (Electrical Engineering). PWS Pub. Co. 1995.

6. Goldberg D. E. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley Professional. 1989.

7. Holland J. H. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. The MIT Press. 1992.

8. Haupt R. L. Practical Genetic Algorithms, Wiley-Interscience. 1997.

9. Wand M., Jones M. Kernel Smoothing. New York : Chapman and Hall. 1995.

А. В. Кузнецов, А. И. Рубан

## АЛГОРИТМЫ МЕТОДА УСРЕДНЕНИЯ КООРДИНАТ ПРИ ПОИСКЕ ГЛАВНЫХ МИНИМУМОВ МНОГОЭКСТРЕМАЛЬНЫХ ФУНКЦИЙ

*Построены алгоритмы поиска заданного количества главных минимумов многоэкстремальных функций многих непрерывных переменных при активном учете ограничений неравенств. В основе алгоритмов лежит разбиение заданной области поиска на подобласти, тяготеющие к требуемым главным минимумам, и последующий поиск в каждой найденной подобласти условного глобального экстремума на основе алгоритмов метода усреднения координат. Разбиение на подобласти производится также на основе алгоритмов усреднения координат путем их последовательных запусков и исключением уже найденных подобластей с помощью дополнительных ограничений неравенств. На численных примерах продемонстрирована эффективность работы алгоритмов.*

*Ключевые слова: глобальная оптимизация, главные минимумы.*

При решении проблемы многокритериальной оптимизации часто основываются на одном из показателей эффективности и минимизируют (или максимизируют) его, а остальные показатели выдерживают не хуже заданных (так появляются дополнительные ограничения неравенства). В ограничения неравенства входят также ресурсные и конструктивные ограничения. При такой оптимизации исследователя интересу-

ет не только положение глобального минимума выбранного показателя эффективности, но и положение близлежащих (по величине минимизируемой функции) минимумов. Указанные минимумы назовем главными.

В работе [1] предложен алгоритм поиска главных минимумов без ограничений неравенств, в [2, с. 139–143] были проведены испытания алгоритма на простейших