A Benchmark of Non-intrusive Parametric Audio Quality Estimation Models for Broadcasting Systems and Web-casting Applications

Martin JAKUBIK, Peter POCTA

Department of Multimedia and Information and Communication Technology, Faculty of Electrical Engineering and Information Technology, University of Zilina, Univerzitna 8215/1, 010 26 Zilina, Slovak Republic

martin.jakubik@feit.uniza.sk, peter.pocta@feit.uniza.sk

DOI: 10.15598/aeee.v19i4.4207

Article history: Received Apr 12, 2021; Revised Jun 24, 2021; Accepted Aug 31, 2021; Published Sep 30, 2021.

Abstract. *Due to the rising usage of various broadcasting systems and web-casting applications, a measurement of audio quality has become an essential task. This paper presents a benchmark of the parametric models for non-intrusive estimation of the audio quality perceived by the end user. The proposed solution is based on machine learning techniques for broadcasting systems and web-casting applications. The main goal of this study is to assess the performance of the non-intrusive parametric models as well as to evaluate a statistical significance of the performance differences between those models. The paper provides a comparison of several models based on the Support Vector Regression, Genetic Programming, Multigene Symbolic Regression, Neural Networks and Random Forest. The obtained results indicate that among the investigated models the most accurate, although not the fastest ones, are the model based on Random Forest (a broadcast scenario) and the SVR-based model (a web-cast scenario). These models represent promising candidates for non-intrusive parametric audio quality assessment in the context of broadcasting systems and web-casting applications.*

Keywords

Artificial Neural Networks, audio quality estimation, broadcast, machine learning, statistical significance, Support Vector Regression, web-cast.

1. Introduction

The coronavirus (SARS-CoV-2) pandemic has led to a sudden and major shift towards online space. People spend more and more time at home, which also leads to an increase in the usage of broadcasting systems and web-casting applications. Therefore, the number of providers of these services is also increasing. However, with the growing number of customers, there is a need to constantly improve these systems and applications because only those providers who will provide the best services will keep these users/customers. In the world of broadcasting and web-casting technology, the quality perceived by the end users plays an important role in increasing customer loyalty to these services. The audio or video quality provided by these broadcasting systems and web-casting applications will be an essential aspect of their acceptance among the general population. The assessment of the audio quality of broadcasting systems and web-casting applications will therefore be a key factor to understand the degree of satisfaction of the end-users in an environment of this kind.

The most accurate approach used for evaluating the quality of speech and audio is a subjective evaluation. The panel of listeners is asked to rate a quality of audio recordings, often using rating scale ranging from an excellent quality to a bad quality [1]. Subsequently, these scores are averaged and a Mean Opinion Score (MOS) is obtained. Although this process is the most accurate, it also has many flaws such as a time demanding nature, lack of repeatability and higher financial costs. For that reason, this approach is inappropriate for real-time applications [2]. Therefore, many objec-

tive quality measurement models have been proposed to estimate MOS values.

Approaches to assess objective audio quality can be separated into two main groups, i.e. intrusive and non-intrusive [3]. Currently, an intrusive technique is used in the most instances. In the case of the intrusive approach, a distortion between the reference and degraded signal that has been processed by a system under test is compared. On the other hand, in situations when a reference signal is not available, e.g. in real-time monitoring systems or wireless communication, the objective quality measurement must be carried out without a reference signal, i.e. by means of a non-intrusive approach [3]. Currently, no non-intrusive parametric audio quality assessment model that focuses on broadcasting systems and web-casting applications is standardized by International Telecommunication Union despite the fact that the non-intrusive models of parametric quality evaluation are standardized for speech [4] and an audio-visual media streaming [5].

Therefore, in this article, we present non-intrusive parametric models of audio quality estimation based on machine learning methods for broadcasting systems and web-casting applications. Subsequently, we compare a performance of these models and evaluate a statistical significance of the corresponding performance differences in order to identify the most effective one for a real deployment.

The remaining of the paper is organized as follows. Section 2. describes a database deployed for a design as well as a benchmark of the investigated models. An experimental methodology is presented in Sec. 3. Section 4. gives details of the investigated models. Section 5. presents the results of the model benchmarks. Section 6. concludes the paper.

2. Dataset

The dataset used for the experiments consists of 3 hours long recordings of uncompressed audio from the Slovak Radio sampled at 48 kHz/16-bit. The dataset was supplemented with the European Broadcasting Union (EBU) Sound Quality Assessment Material (SQUAM) music database. The version of the database used is described in the EBU Tech 3253 [9]. A total of 27 diverse types of signals represented by stereo music samples of duration of 10–15 s, reflecting variety of audio signals transmitted to the public by broadcasting systems and web-casting applications, are included in the database. Taking into account [8], [10] and [11] and similarly as in the case of [6] and [7], we have chosen the following input parameters for broadcasting systems, i.e. a type of codec, type of audio signal and bit rate. The used codecs and their bit rates can be found in [6], see Tab. 1 for more details. To assess a perceived audio quality, we used the Perceptual Objective Listening Quality Analysis (POLQA) Music V2 model, see [8] for more information.

The broadcasting sub-database contains 1,080 MOS-LQO (Mean Opinion Score - Listening-only Quality Objective) values, a combination of the different codecs, bit rates and signal types. For web-casting applications we have considered other degradations parameters on the top of them, based on [12], [13] and [14]. So, we have added a stalling and an initial delay to the above mentioned degradation parameters (the type of signal, type of codec and bit rate) as the final audio quality is also influenced by these two occurrences, which finally reduce the MOS-LQO value. In general, the effect of the initial delay and the stalling on perceived quality depends only on their duration. In order to reflect real-world conditions, we have used the real-world measurements presented in [12] and [13] to select the values of initial delay and stalling to be deployed to create the database, see [6] for the selected values.

Since the POLQA Music was not trained for degradations induced by the initial delay and stalling, their impact on a quality experienced by the end user was evaluated by the model defined in [14]. As the coding impairment and the initial delay and stalling impairments are different in nature, i.e. their frequency and time domain, we can apply the additivity concept that comes from the E-model [4] to get the combined impact of these impairments on a quality experienced by the end user. Overall, along with additional degradations we have 17,280 MOS-LQO values that define the webcasting environment, i.e. a web-cast sub-database. Both sub-databases were divided into a training part and testing part at a ratio of 80:20. Further details about how the database was built can be found in [6].

3. Experimental Methodology

In this work, we have benchmarked non-intrusive parametric audio quality estimation models for broadcasting systems (a diagram of this model type is depicted in Fig. 1) and web-casting applications (a diagram of this model type is depicted in Fig. 2) based on several machine learning methods, namely Support Vector Regression (SVR), Artificial Neural Network (ANN), Random Forest (RF), Genetic Programming (GP) and Multigene Symbolic Regression (MSR) in order to identify the most effective one for a real deployment. It is worth noting here that the database described in Sec. 2. was deployed to train and test/benchmark the abovementioned models. The MOS-LQE (Mean Opinion Score - Listening-only Quality Estimated) val-

ues obtained by these models were compared with the MOS-LQO values.

The efficiency of the parametric estimation models was quantified in terms of the Pearson Correlation Coefficient (PCC) and the respective Root Mean Square Error (RMSE) widely used in this context. To provide the information on the significance of the differences between the presented PCC and RMSE values for the above mentioned models, the corresponding statistical significance tests, were performed, see [15] for more detail. It is worth noting here that the investigated models formulate an estimate of audio quality as a regression problem, to be solved by the particular machine learning technique, to find a mapping between the audio features and quality score. The corresponding process is briefly described below.

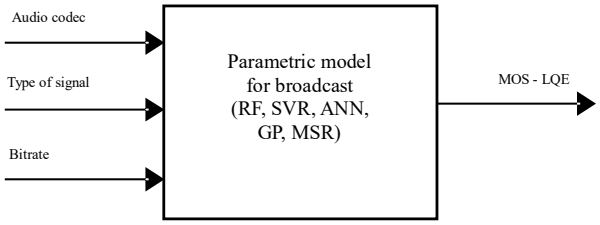

Fig. 1: Diagram of the proposed parametric prediction model for broadcasting systems.

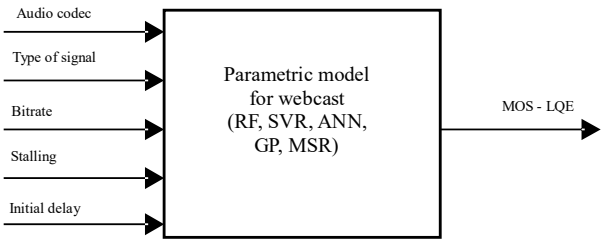

Fig. 2: Diagram of the proposed parametric prediction model for web-casting applications.

The SVR [16], ANN [17], RF [18], MSR [19], [20] and [21] and GP [22], [23] and [24] are methods that use a supervised learning. In supervised machine learning the training data would consist of inputs (X) combined with the right outputs (Y). The algorithm will look for patterns in the data during a training that correspond with the desired outputs. The aim is to approximate the mapping function so well that the output variables (Y) for that data can be predicted when you have new input data (X). So, a supervised learning algorithm can be described in its most basic form as follows:

$$Y = f(x). \tag{1}$$

In this research, the machine learning techniques were deployed because they have shown to be successful for the related tasks according to the literature, see [25], [26], and [27] for more detail.

4. Description of Investigated Models

Our aim was to design non-intrusive parametric models based on the machine learning approach. We will briefly discuss in the following subsections the SVR and ANN based models as they were not published yet. A description of the other parametric audio quality estimation models based on Random Forest, Genetic Programming and Multigene Symbolic Regression approaches involved in this benchmark can be found in [6], [7], and [28], respectively.

4.1. Support Vector Regression

The SVR [16] technique is based on the Support Vector Machine (SVM) firstly introduced by Vapnik at the end of 20th century [29] and is firmly anchored in the statistical learning theory. The basic idea is to map an input data to a large dimension space using a non-linear mapping and then a problem of linear regression is obtained in that space. The SVR approach has indisputable advantages, such as an ability to capture non-linear data dependencies or a simplicity of the model created, since the whole solution is represented only by a subset of support vectors. In addition to these positive features, however, a usage of the SVR approach poses a problem of selecting the right internal parameters. In order to use SVR to solve a problem, it is necessary to first define several internal parameters. Choosing the right values dramatically affects the performance of this method. In general, the parameters affect the model's ability to generalize and hence the accuracy of the estimation. In the context of SVR, there are two groups of internal parameters. The first group includes the parameters of the SVR algorithm itself, e.g. parameter C and ε. Another category/group is represented by the parameters of the selected kernel function, such as the gamma parameter determining a width of the Gauss kernel. In our work we dealt with optimizing these parameters to obtain the most accurate estimations and the best models. The overall goal of the SVR algorithm is to find a function $f(x)$ that minimizes the model error with respect to its parameters.

In our case, ε was set to 0.15. In essence, the SVR focuses upon the small subset of examples that are important to estimate the quality. We further noticed that as C increases, a number of SVs also increases. The reason behind it is rather simple as C is a penalty for the errors and is used to weigh the outliers. Obvi-

ously, as C is increased, the system tends to put large weight on the outliers. When overfitting occurs the SVR will choose a very large number of data points as SVs in order to achieve a good performance [30].

4.2. ANN

ANN models are effective non-linear modelling methods that simulate a human brain. Models of ANN correspond to biological neural networks. A complex network mapping input variables to output variables can be generated by an ANN, being able to estimate non-linear functions. A structure of each ANN consists of an input layer of neurons/nodes, at least one hidden layer of neurons/nodes and a final layer of output neurons/nodes [17]. Artificial Neural Network with one hidden layer is considered a shallow structure model while Artificial Neural Network with more than one hidden layer is a multi-layer Artificial Neural Network. In our case, both types of the ANN structure were used. When it comes to a design of ANN-based model, a key task is to define input variables and an optimal configuration of the network in order to accurately generate a desired output. A number of hidden neurons in a multi-layer neural network and number of nodes included in them were selected empirically in the case of this experiment.

5. Results

In this section, we will first present a performance evaluation of two newly designed models, i.e. ANN and SVR based models, which was not yet published. Secondly, we will benchmark all the investigated models in order to identify the most effective one for a real deployment.

For the ANN-based models, an experiment was performed 10 times and the best results are reported. The MOS-LQE values provided by the parametric ANN-based estimation models were compared with the MOS-LQO values of the test sets of the corresponding sub-datasets. The correlation obtained by the shallow NN and ANN calculated over all the test conditions reached 0.8341 and 0.8606 for the broadcast scenario and 0.9657 and 0.9749 for the web-cast scenario.

Moreover, the obtained RMSE values are reported in Tab. 1 and Tab. 2 for the broadcast and web-cast scenario, respectively. Figure 3 and Fig. 5 compare the MOS-LQO values and the MOS-LQE values obtained by the designed models for the broadcast scenario. On the other hand, Fig. 4 and Fig. 6 compare the MOS-LQO values and the MOS-LQE values for the web-cast scenario.

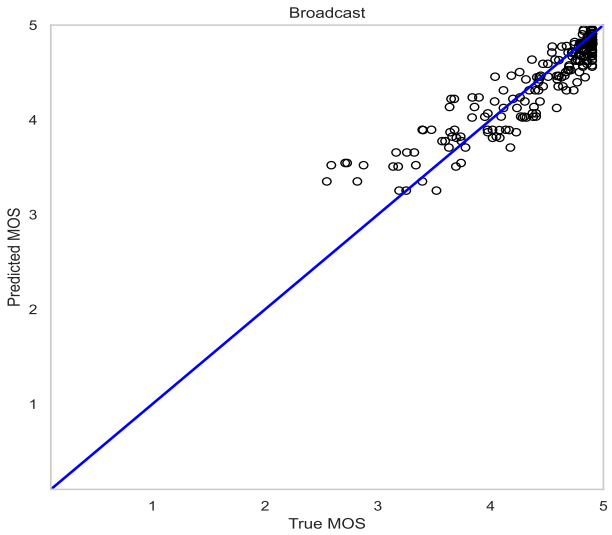

Fig. 3: Scatter plots of the MOS-LQO values versus the MOS-LQE values obtained by the shallow NN approach for the broadcast scenario.

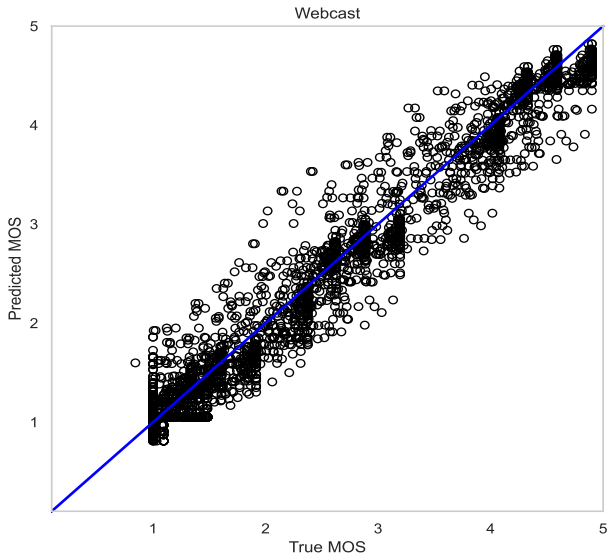

Fig. 4: Scatter plots of the MOS-LQO values versus the MOS-LQE values obtained by the shallow NN approach for the web-cast scenario.

Similarly as in the Artificial Neural Network case, we proceeded with the Support Vector Regression method. Thus, 10 experiments were conducted for each sub-database and the best results were noted. Again, we compared the MOS-LQE values obtained by the designed model with the MOS-LQO values. The results show a very good success rate by using the SVR approach, as the Pearson's correlation coefficient and RMSE reached 0.9267 and 0.2348 for the broadcast conditions, and 0.9889 and 0.1946 for the web-cast conditions, respectively. For a visual comparison, scatter plots for the models based on the SVR approach are shown in Fig. 7 and Fig. 8.

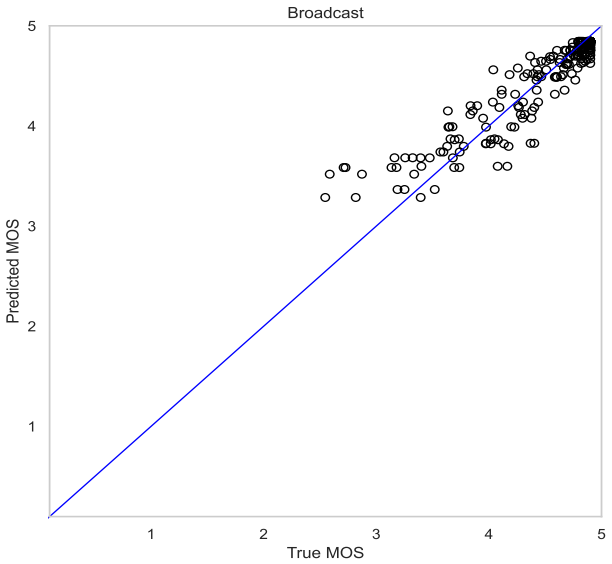

Fig. 5: Scatter plots of the MOS-LQO values versus the MOS-LQE values obtained by the ANN approach for the broadcast scenario.

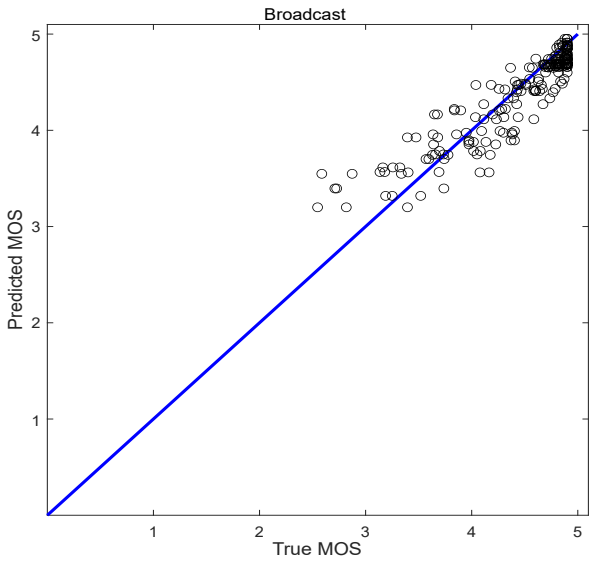

Fig. 7: Scatter plots of the MOS-LQO values versus the MOS-LQE values obtained by the SVR approach for the broadcast scenario.

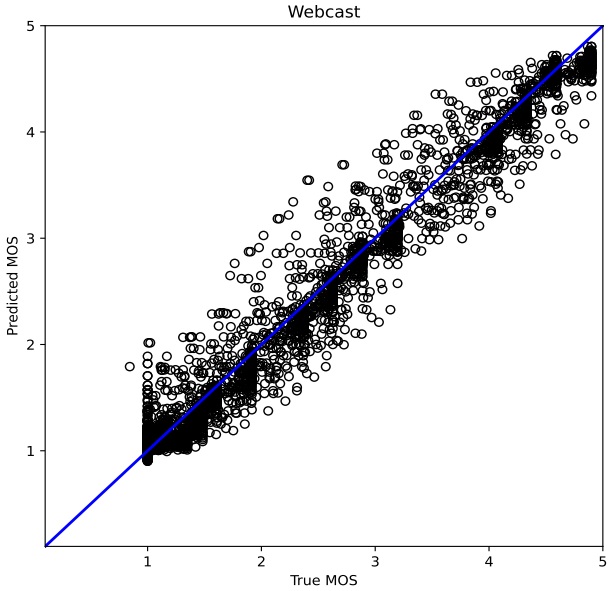

Fig. 6: Scatter plots of the MOS-LQO values versus the MOS-LQE values obtained by the ANN approach for the web-cast scenario.

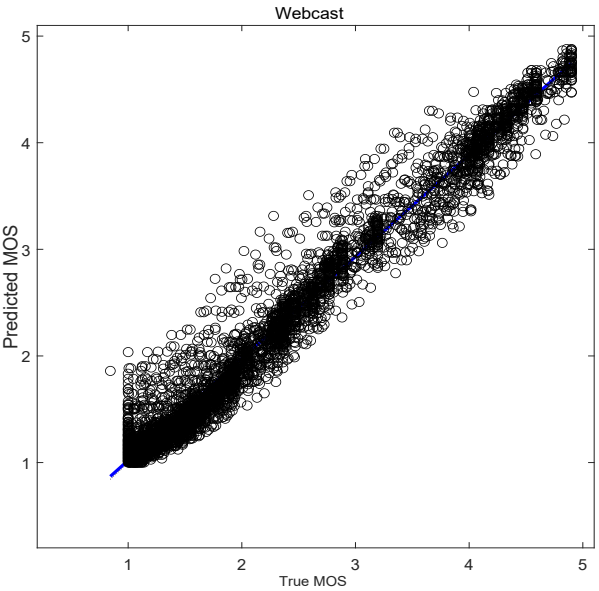

Fig. 8: Scatter plots of the MOS-LQO values versus the MOS-LQE values obtained by the SVR approach for the web-cast scenario.

As it can be noted in the scatter plots, there are some outliers in the both investigated scenarios, i.e. broadcast and web-cast. The outliers obtained for the SVR approach are dominantly populated by the music signals coded by MP2 (overestimation), AAC-LC (underestimation) and HE-AACv2 (underestimation) codecs, at the lower deployed bit rates. When it comes to the outliers for the speech signals, they are solely represented by MP3 codec (underestimation) operating at the lower deployed bit rates. In the case of the ANNs and shallow NNs models, the outliers are prevalently represented by the music signals coded by AAC-LC (underestimation), HE-AACv2 (over- and under-estimation) and MP2 (overestimation) codecs, at the lower deployed bit rates. It should be noted here that the same set of the codecs is reported for both approaches in this case. For the speech signals, the outliers are dominantly covered by the MP3 (underestimation), AAC-LC (underestimation), HE-AACv2 and MP2 (both over- and under- estimation) codecs. It is worth noting that when it comes to the web-cast scenario, which has involved more training data, the estimates are naturally more consistent in both cases, i.e. the SVR and ANN models.

The obtained PCC and RMSE values for all the investigated models and both kind of the scenarios (broadcast and web-cast) are given in Tab. 1 and Tab. 2. The Pearson correlation coefficient calculated for all the investigated models varies between 0.8341 and 0.9411 for the broadcast scenario and 0.9657 and 0.9889 for the web-cast scenario. As it can be seen from Tab. 1, the highest value of the correlation for the broadcast was reached by the model based on the RF approach. On the other hand, when it comes to the web-cast scenario, it was the model based on the SVR approach, see Tab. 2 for more detail. Both models have also achieved the lowest RMSE value.

Tab. 1: Pearson correlation coefficients and root mean square errors obtained for the investigated models for the broadcast scenario. The best performing model in terms of both investigated metrics, i.e. the PCC and RMSE, is highlighted.

Model	PCC	RMSE
Genetic Programming	0.8988	0.2458
Multigene Symbolic Regression	0.8502	0.2126
Random Forest	**0.9411**	**0.1951**
Support Vector Regression	0.9267	0.2348
Shallow Neural Network	0.8341	0.2237
Artificial Neural Network	0.8606	0.2050

Tab. 2: Pearson correlation coefficients and root mean square errors obtained for the investigated models for the web-cast scenario. The best performing model in terms of both investigated metrics, i.e. the PCC and RMSE, is highlighted.

Model	PCC	RMSE
Genetic Programming	0.9708	0.3187
Multigene Symbolic Regression	0.9748	0.2015
Random Forest	0.9854	0.2192
Support Vector Regression	**0.9889**	**0.1946**
Shallow Neural Network	0.9657	0.2348
Artificial Neural Network	0.9749	0.2009

To specify the significance of the differences between the presented Pearson correlation coefficient and Root Mean Square Error values for the broadcast and web-cast scenario, the corresponding statistical significance tests, were performed, see [15] for more detail.

Tab. 3: Results of statistical significance tests for the Pearson Correlation Coefficients and Root Mean Square Errors for the broadcast scenario. Note: "1" indicates that the difference is statistically significant. "0" indicates that the difference is not statistically significant.

Broadcast		
	PCC	RMSE
Genetic Programming	1	1
Multigene Symbolic Regression	1	0
Support Vector Regression	0	1
Shallow Neural Network	1	0
Artificial Neural Network	1	0

The results of such tests for the broadcast scenario are displayed in Tab. 3, where we compare the best performing model, i.e. the RF-based model, in terms of the PCC and RMSE with the other models. It should be noted here that a value of 1 implies that there is a statistically significant difference and a value of 0 indicates that there is no statistically significant difference. Table 3 shows that most of the differences are statistically significant. It means that the models are statistically different in such cases. Regarding the web-cast scenario, the results presented in Tab. 4 show that mostly all the differences are statistically significant.

Tab. 4: Results of statistical significance tests for the Pearson Correlation Coefficients and Root Mean Square Errors for the web-cast scenario. Note: "1" indicates that the difference is statistically significant. "0" indicates that the difference is not statistically significant.

Web-cast		
	PCC	RMSE
Genetic Programming	1	1
Multigene Symbolic Regression	1	0
Random Forest	1	1
Shallow Neural Network	1	1
Artificial Neural Network	1	0

Moreover, we have compared a computational load, i.e. a time taken by the trained model to make a quality estimation for the corresponding set of the input parameters, of the models investigated in this study. The computational load varies between 34.56 milliseconds and 43.04 milliseconds for the broadcast scenario and 137.59 and 534.03 milliseconds for the web-cast scenario. All the computational load values, including the ones obtained for the GP, MSR and RF models, are listed in Tab. 5. The results show that the models based on the Genetic Programming approach have achieved the lowest computational load in both cases (broadcast and web-cast). On the other hand, the highest computational load was obtained for the model based on the Random Forest approach for the broadcast scenario and the Support Vector Regression approach for the web-cast scenario, interestingly the most accurate ones. It should be noted here that all the experiments were carried out on a 64-bit quad-core processor based on the Kaby Lake H Architecture, Intel i7-7700HQ 2.8 GHz.

Tab. 5: Computational load of the investigated models for the broadcast as well as the web-cast scenario.

Computational load	Broadcast (ms)	Web-cast (ms)
Genetic Programming	17.01	27.02
Multigene Symbolic Regression	16.90	27.95
Random Forest	65.54	285.3
Support Vector Regression	43.04	534.03
Artificial Neural Network	40.36	220.90
Shallow Neural Network	34.56	137.59

6. Conclusion

In this paper, we presented a design and performance evaluation of the parametric models for non-intrusive estimation of the audio quality based on machine learning techniques, namely the ANN and SVR, for broadcasting systems and web-casting applications. The main goal of this study was to benchmark the proposed models as well as to evaluate the statistical significance of the performance differences between them. In addition to the presented models, we also compared models based on Genetic Programming, Random Forest and Multigene Symbolic Regression.

By comparing all the results we can say, that the best performance in terms of broadcasting systems, was achieved by a model based on the Random Forest approach. This model achieved the highest correlation and at the same time the lowest RMSE value. On the other hand, when it comes to the web-casting scenario, the best performance was achieved for the model based on the Support Vector Regression approach. From the computational load perspective, the best results were achieved for the models based on the Genetic Programming approach. However, computational load is a secondary component here and can ultimately be reduced by upgrading to a more powerful machine and/or using an optimized implementation.

We conclude that the best performing models, i.e. the RF-based model (the broadcast scenario) and the SVR-based model (the web-cast scenario), are the most effective ones for a real deployment when it comes to a quality planning and monitoring of broadcasting systems and web-casting applications. So, the models are intended to help broadcasters and web-casters to identify the best configuration of their systems and services in terms of quality experienced by the end user.

Author Contributions

Both M.J. and P.P. developed the theoretical formalism, performed the analytic calculations and performed the machine learning simulations. Both M.J. and P.P. authors contributed to the final version of the manuscript.

References

[1] BEERENDS, J. G., C. SCHMIDMER, J. BERGER, M. OBERMANN, R. ULLMANN, J. POMY and M. KEYHL. Perceptual Objective Listening Quality Assessment (POLQA), The Third Generation ITU-T Standard for End-to-End Speech Quality Measurement Part I–Temporal Alignment. *Journal of the Audio Engineering Society*. 2013, vol. 61, iss. 6, pp. 366–384. ISSN 1549-4950.

[2] ZHENG, J., M. ZHU and Y. SONG. On objective assessment of audio quality – A review. In: *2012 International Conference on Audio, Language and Image Processing*. Shanghai: IEEE, 2012, pp. 777–782. ISBN 978-1-4673-0174-9. DOI: 10.1109/ICALIP.2012.6376719.

[3] RIX, A. W., J. G. BEERENDS, D.-S. KROON and O. GHITZA. Objective Assessment of Speech and Audio Quality—Technology and Applications. *IEEE Transactions on Audio, Speech, and Language Processing*. 2006, vol. 14, iss. 6, pp. 1890–1901. ISSN 1558-7924. DOI: 10.1109/TASL.2006.883260.

[4] ITU-T G.107. *The E-model: a computational model for use in transmission planning*. Geneva: ITU-T, 2015.

[5] ITU-T P.1201. *Parametric non-intrusive assessment of audiovisual media streaming quality – Lower resolution application area*. Geneva: ITU-T, 2012.

[6] JAKUBIK, M. and P. POCTA. Non-Intrusive Parametric Audio Quality Estimation Models for Broadcasting Systems and Web-Casting Applications Based on Random Forest. *Advances in Electrical and Electronic Engineering*. 2020, vol. 18, iss. 4, pp. 235–241. ISSN 1804-3119. DOI: 10.15598/aeee.v18i4.3890.

[7] JAKUBIK, M. and P. POCTA. Parametric audio quality estimation models for broadcasting systems and web-casting applications based on the Genetic Programming. In: *2020 18th International Conference on Emerging eLearning Technologies and Applications (ICETA)*. Kosice: IEEE, 2020, pp. 219–225. ISBN 978-1-6654-2226-0. DOI: 10.1109/ICETA51985.2020.9379251.

[8] POCTA, P. and J. G. BEERENDS. Subjective and Objective Assessment of Perceived Audio Quality of Current Digital Audio Broadcasting Systems and Web-Casting Applications. *IEEE Transactions on Broadcasting*. 2015, vol. 61, iss. 3, pp. 407–415. ISSN 1557-9611. DOI: 10.1109/TBC.2015.2424373.

[9] EBU TECH 3253. *Sound Quality Assessment Material recordings for subjective tests*. Geneva: EBU, 2008.

[10] LEE, S., Y.-T. LEE, J. SEO, M.-S. BAEK, C.-H. LIM and H. PARK. An Audio Quality Evaluation of Commercial Digital Radio Systems. *IEEE Transactions on Broadcasting*. 2011,

vol. 57, iss. 3, pp. 629–635. ISSN 1557-9611. DOI: 10.1109/TBC.2011.2152910.

[11] BERG, J., C. BUSTAD, L. JONSSON, L. MOSSBERG and D. NYBERG. Perceived Audio Quality of Realistic FM and DAB+ Radio Broadcasting Systems. *Journal of the Audio Engineering Society*. 2013, vol. 61, iss. 10, pp. 755–777. ISSN 1549-4950.

[12] SCHWIND, A., F. WAMSER, T. GENSLER, P. TRAN-GIA, M. SEUFERT and P. CASAS. Streaming Characteristics of Spotify Sessions. In: *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*. Cagliari: IEEE, 2018, pp. 1–6. ISBN 978-1-5386-2605-4. DOI: 10.1109/QoMEX.2018.8463372.

[13] SCHWIND, A., L. HABERZETTL, F. WAMSER and T.HOSSFELD. QoE Analysis of Spotify Audio Streaming and App Browsing. In: *Proceedings of the 4th Internet-QoE Workshop on QoE-based Analysis and Management of Data Communication Networks (Internet-QoE'19)*. Los Cabos: ACM, 2019, pp. 25–30. ISBN 978-1-4503-6927-5. DOI: 10.1145/3349611.3355546.

[14] SACKL, A., S. EGGER and R. SCHATZ. Where's the music? comparing the QoE impact of temporal impairments between music and video streaming. In: *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*. Klagenfurt am Worthersee: IEEE, 2013, pp. 64–69. ISBN 978-1-4799-0738-0. DOI: 10.1109/QoMEX.2013.6603212.

[15] ITU-T P.1401. *Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models*. Geneva: ITU-T, 2012.

[16] SMOLA, A. J. and B. SCHOLKOPF. A tutorial on support vector regression. *Statistics and Computing*. 2004, vol. 14, iss. 1, pp. 199–222. ISSN 1573-1375. DOI: 10.1023/B:STCO.0000035301.49549.88.

[17] WANG, S.-C. Artificial Neural Network. *Interdisciplinary Computing in Java Programming*. 2003, vol. 743, iss. 1, pp. 81–100. ISSN 1573-1375. DOI: 10.1007/978-1-4615-0377-4_5.

[18] BREIMAN, L. Random Forests. *Machine Learning*. 2001, vol. 45, iss. 1, pp. 5–32. ISSN 1573-0565. DOI: 10.1023/A:1010933404324.

[19] SEARSON, D. P., D. E. LEAHY, M. J. WILLIS. GPTIPS: An Open Source Genetic Programming Toolbox For Multigene Symbolic Regression. In: *International MultiConference of Engineers and Computer Scientists (IMECS)*. Hong Kong: Newswood Ltd., 2010, pp. 77–80. ISBN 978-988-17012-8-2.

[20] SEARSON, D. P. *Handbook of Genetic Programming Applications: GPTIPS 2: An open-source software platform for symbolic data mining*. 1st ed. Berlin: Springer, 2015. ISBN 978-3-319-20883-1.

[21] RAHDARI, F., M. EFTEKHARI and D. R. MOUSAVI. A two-level multi-gene genetic programming model for speech quality prediction in Voice over Internet Protocol systems. *Computers & Electrical Engineering*. 2016, vol. 49, iss. 1, pp. 9–24. ISSN 0045-7906. DOI: 10.1016/j.compeleceng.2015.10.008.

[22] VERGILIO, S. R. and A. POZO. A grammer-guided Genetic Programming framework configured for data mining and softtware testing. *International Journal of Software Engineering and Knowledge Engineering*. 2006, vol. 16, iss. 2, pp. 245–267. ISSN 0218-1940. DOI: 10.1142/S0218194006002781.

[23] BANZHAF, W., P. NORDIN, R. E. KELLER and F. D. FRANCONE. *Genetic Programming: An Introduction: The Morgan Kaufmann Series in Artificial Intelligence*. 1st ed. Amsterdam: Elsevier Science, 1998. ISBN 978-1-558-60510-7.

[24] SILVA, S. and J. ALMEIDA. GPLAB – A Genetic Programming Toolbox for MATLAB. In: *Proceedings of the Nordic MATLAB Conference (NMC)*. Copenhagen: MathWorks, 2003, pp. 273–278. ISBN 87-989426-0-3.

[25] NARWARIA, M., W. LIN, I. V. MCLOUGHLIN, S. EMMANUEL and C. L. TIEN. Non-intrusive Speech Quality Assessment with Support Vector Regression. In: *Advances in Multimedia Modeling: 16th International Multimedia Modeling Conference (MMM)*. Heidelberg: Springer, 2003, pp. 325–335. ISBN 978-3-642-11301-7. DOI: 10.1007/978-3-642-11301-7_34.

[26] MIZDOS, T., M. BARKOWSKY, M. UHRINA and P. POCTA. Linking Bitstream Information to QoE: A Study on Still Images Using HEVC Intra Coding. *Advances in Electrical and Electronic Engineering*. 2019, vol. 17, iss. 4, pp. 436–445. ISSN 1804-3119. DOI: 10.15598/aeee.v17i4.3625.

[27] AVILA, A. R., H. GAMPER, C. REDDY, R. CUTLER, I. TASHEV and J. GEHRKE. Non-intrusive speech quality assessment using neural networks. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Brighton: IEEE, 2019, pp. 631–635. ISBN 978-1-4799-8131-1. DOI: 10.1109/ICASSP.2019.8683175.

[28] JAKUBIK, M. and P. POCTA. Estimating the Perceived Audio Quality Based on Multi-gene Symbolic Regression for Broadcasting Systems and Web-Casting Applications. In: *31st International Conference Radioelektronika (RADIOELEKTRONIKA)*. Brno: IEEE, 2021, pp. 1–5. ISBN 978-1-6654-1474-6. DOI: 10.1109/RADIOELEKTRONIKA52220.2021.9420201.

[29] CORTES, C. and V. VAPNIK. Support-vector networks. *Machine Learning*. 1995, vol. 20, iss. 1, pp. 273–297. ISSN 1573-0565. DOI: 10.1007/BF00994018.

[30] NARWARIA, M., W. LIN, I. V. MCLOUGHLIN, S. EMMANUEL and L.-T. CHINA. Nonintrusive Quality Assessment of Noise Suppressed Speech With Mel-Filtered Energies and Support Vector Regression. *IEEE Transactions on Audio, Speech, and Language Processing*. 2012, vol. 20, iss. 4, pp. 1217–1232. ISSN 1558-7924. DOI: 10.1109/TASL.2011.2174223.

About Authors

Martin JAKUBIK was born in Ruzomberok, Slovak Republic. He received his M.Sc. in Telecommunications from Faculty of Electrical Engineering, University of Zilina, Slovak Republic in 2018. He is currently a Ph.D. student at the Department of Multimedia and Information-Communication Technology of the University of Zilina. His current research interests include speech and audio quality estimation using machine learning methods.

Peter POCTA was born in 1981. He received his M.S. and Ph.D. degrees from University of Zilina, Faculty of Electrical Engineering, Slovak Republic in 2004 and 2007, respectively. He is currently a Full Professor at the Department of Multimedia and Information-Communication Technology of the University of Zilina and is involved with International Standardization through the European Telecommunications Standards Institute Technical Committee Speech and Multimedia Transmission Quality (ETSI TC STQ) as well as ITU-T SG12. His research interests include speech, audio, video and audiovisual quality assessment, speech intelligibility, multimedia communication and QoE management. He has published over 60 peer-reviewed papers in international journals and conferences including Acta Acustica united with Acustica, Applied Acoustics (Elsevier), Speech Communication (Elsevier), IEEE Transactions on Broadcasting, Measurement of Speech, Audio and Video Transmission Quality In Telecommunication Networks (MESAQIN) and Quality of Multimedia Experience (QoMEX) conferences. He serves as an external reviewer for the Journal of Systems and Software (Elsevier), IEEE Transactions on Multimedia, Multimedia Systems (Springer), Speech Communication (Elsevier), Telecommunication Systems (Springer), Quality and User Experience (Springer) and IEEE/ACM Transactions on Audio, Speech and Language Processing and several conferences in area of multimedia quality and communication networks, e.g. QoMEX, QCMan, etc.