

NON-INTRUSIVE PARAMETRIC AUDIO QUALITY ESTIMATION MODELS FOR BROADCASTING SYSTEMS AND WEB-CASTING APPLICATIONS BASED ON RANDOM FOREST

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/aeec.v18i4.3890

Abstract. *Objective quality assessment models have been used more and more in recent years to assess or monitor speech and audio quality in many multimedia and audio processing systems. These methods offer a clear and repeatable way to evaluate a customer experience by measuring perceived quality on a subjective scale, which is easily understood, such as a quality rating scale, ranging from excellent quality to a low quality. Subsequently, the aim of service providers is to offer reliable services providing the end-user/customer with the best possible quality in the context of the current network conditions to avoid a customer churn. This paper presents a design and performance evaluation of parametric models estimating the audio quality experienced by the end user of broadcasting systems and web-casting applications. The Random Forest (RF) algorithm is used to design non-intrusive parametric models, establishing the relationship between the feature description and the perceived quality scores. For this, the broadcast and web-cast sub-databases were created, where the web-cast sub-database includes 17,280 degraded samples and the broadcast sub-database contains 1,080 degraded samples obtained from the Slovak Radio. The results reported for the proposed parametric audio quality models have validated Random Forest as a powerful technique that provides a good efficiency in terms of Pearson Correlation Coefficient (PCC) and Root Mean Squared Error (RMSE).*

Keywords

Audio quality estimation, broadcast, initial delay, Random Forest, stalling, web-cast.

1. Introduction

The substantial development in broadcasting and networking technologies in recent decades reflects the range of digital audio transmission systems, such as Digital Audio Broadcasting (DAB) [1] and [2], Digital Radio Mondiale (DRM) [3] and [4], etc., or web-casting applications using TCP as a transport protocol [5]. On-demand streaming media such as Spotify, Deezer, iTunes, Amazon Music or YouTube Music are among the most common audio streaming services [6]. The streaming service, as well as the broadcasting systems, use audio codecs to minimize bandwidth to be used for the corresponding data transmission. Due to the robustness and reliability of the channel coding of today's digital audio transmission systems, the audio codec is still the key source of quality degradation in this case. The same is true for the web-casting applications as they dominantly deploy the TCP on a transport layer [7]. Given that most of today's web-casting applications are based on HTTP streaming [8], additional quality influencing factors, in this case, are an initial delay and a stalling [5]. The creation of effective quality monitoring tools operating in real-time that can measure the audio quality experienced by the end user is therefore important for the success of any audio service or application.

In order to get a clear picture about the deployment of codecs in the context of digital audio broadcasting systems and web-casting applications, we compiled a list of the currently most widely used codecs and their bit rates [9], which is presented in Tab. 1. As it can be clearly seen from this table, MP2 and HE-AACv2 codecs are mostly used when it comes to broadcasting systems. On the other hand, the most widely used codecs for web-casting applications are MP3, Ogg Vor-

bis, AAC-LC, Opus but also lossless codecs such as ALAC and FLAC.

Tab. 1: List of codecs that are used in digital audio broadcasting systems and web-casting applications (Adopted from [7] and updated).

Codec	Bit rate (kb·s ⁻¹)	Applications
MP2	64, 80, 128, 160, 192	DAB
HE-AACv2	24, 32, 40, 48, 56, 80, 96, 112, 128	DAB+, DRM, DRM+, DMB, Web streaming, Deezer, Spotify
MP3	64, 80, 128, 160, 192, 256, 320	Web streaming, Deezer, iTunes
Ogg Vorbis	64, 96, 160, 320	Web streaming, Spotify
AAC-LC	Apple: 96, 128, 256, 320	Web streaming, Wimp, ISDB, iTunes, Spotify
	Nero: 32, 96, 128, 196, 256	
Opus	32, 64, 96, 128, 320	Web streaming

The evaluation of audio quality can be performed from either a subjective or an objective perspective [10]. A subjective listening test is a common way of determining the quality of the audio. Subjective methods that are based on the empirical listening measures defined under international guidelines, e.g. [11] and [12], are more effective but also more arduous and time-consuming. On the other hand, methods for objective assessment of the audio quality are more convenient. To be more precise, objective perceptual audio models are used to reliably and rapidly predict the MOS (Mean Opinion Score) values, i.e., scores representing a quality perceived by the end user.

Generally, objective quality assessment models deploy two approaches [10]:

- intrusive,
- non-intrusive.

Calculating the perceptually weighted difference between the reference and degraded signals is the basic concept of intrusive quality metrics. Intrusive methods are considered more accurate as they provide a higher correlation with subjective evaluations. Intrusive designs in the sense of audio quality, e.g., PEAQ [13], PEMO-Q [14], ViSQOLAudio [15] or POLQA Music [7], expect a qualitative evaluation of the degraded signal distortion by contrasting the degraded signal with the initial/reference signal. So, intrusive measures require the presence of the original signal that is typically not available in continuous quality monitoring.

Moreover, despite the fact that intrusive methods [16] are based on the very hard to realise time synchronization of the two signals, they are still a more reliable counterpart with non-intrusive methods for objective quality evaluation of audio [16]. Non-intrusive

approaches depend on the receiving side processing of the audible signal, without exposure to the original audio signal on the broadcast/web-cast side, which leads to some difficulties in detecting distortion caused by communications networks, i.e., loss of packets. Non-intrusive approaches may be categorized as either signal-based or parametric algorithms [16].

Nevertheless, there is, to the best of our knowledge, currently no non-intrusive parametric model for the audio quality estimation that focuses on the broadcasting systems and web-casting applications despite the fact that these models are already available for speech transmission, i.e., E-model [17], and a broadcast audio contribution over IP [18] and [19] or an audio-visual media streaming [20]. Hence in this paper, we propose non-intrusive parametric audio quality estimation models based on machine learning for the broadcasting systems and web-casting applications. The proposed models formulate an estimate of audio quality as a regression problem and use the RF approach to find a mapping between the audio features and quality score.

The paper is organized as follows. In Sec. 2, we explain an experimental methodology and models input parameters for broadcasting systems and web-casting applications. Section 3. specifies how the database was built to be used to train and validate the proposed models. We then present our proposed non-intrusive parametric quality estimation models, training, and testing phase details in Sec. 4. Section 5. describes the performance evaluation results of the proposed models. Finally, in Sec. 6., we conclude the paper and discuss possible future research.

2. Methodology

In this work, we used Random Forest (RF) as a machine learning technique. Random Forest is a part of the supervised learning methods family [21]. The ecosystem it creates is an ensemble of decision trees [22] and is usually fitted with the bagging cycle. The fundamental theory of bagging methodology is that a combination of learning types increases the combined performance. In that, the size of the forest and the variation between the trees' outputs are minimized. The forest's prediction is the average prediction from individual trees [23]. By running the algorithm on a broad training sample and then ranking these scores against new results, the classification is accomplished. The main aim, of course, is to consider the numerical or logical relationship between the input parameters and performance during the training process.

2.1. Models Input Parameters

We concentrated here on input parameters that could reduce the audio quality of broadcasting systems. As mentioned above, the principal degradation parameter, in this case, is an audio codec. Based on [7], [9] and [24], we also considered the need to take into account a type of signal and bit rate. The corresponding parameters have been used along with the MOS values for training by the Random Forest approach, as illustrated in Fig. 1. The same degradation parameters are used for web-casting applications, see Fig. 2 for more detail, but certain parameters of the application layer have also to be taken into consideration [25], [26], [27], [28] and [29]. An initial delay in audio reproduction is the first of them. The initial delay often occurs since a certain amount of data must be transferred to a receiving side before decoding and playing are going to be initiated. The minimum value for the initial delay depends on the bit rate and encoder settings [5]. Another input degradation parameter that has a significant effect on the perceived audio quality in web-casting applications is stalling [5]. It occurs when actual network throughput is lower than an audio bit rate required by the corresponding streaming service, and a buffer is therefore drained. The effect of the initial delay and stalling on the perceived output generally depends only on its duration.

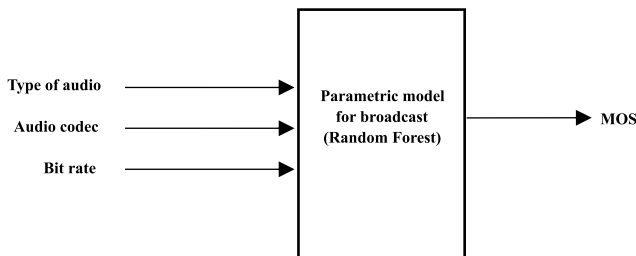


Fig. 1: Diagram of the proposed parametric prediction model for broadcasting systems.

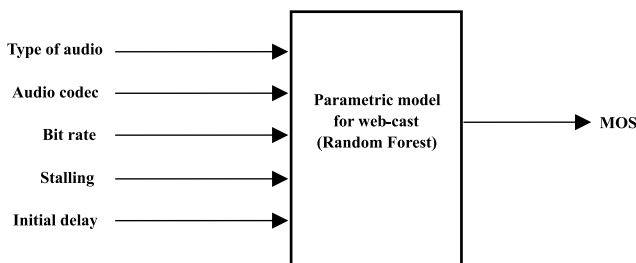


Fig. 2: Diagram of the proposed parametric prediction model for web-casting applications.

To sum up, the following input parameters have been considered in a design phase of non-intrusive parametric audio quality estimation models for broadcasting systems and web-casting applications:

- Broadcast – the type of audio codec, type of signal, bit rate.
- Web-cast – the type of signal, type of audio codec, bit rate, stalling, initial delay.

3. Database

Before creating the database, it was necessary to specify all the types of sounds that we had to focus on. We sought to reach as large audio signal spectrum as possible. The goal was to create the largest available archive of specific recordings that would represent the audio variety transmitted to the listener over broadcasting systems and web-casting applications. To do so, we obtained 3 hours long uncompressed studio recording from Slovak Radio to create a dataset that will be used for training and testing of the designed estimation models. However, it did not contain all the pre-selected audio signals like classical music, a spoken word in a foreign language, a track featuring only one musical instrument, etc. For that reason, we have added a part of the respectable EBU SQUAM database that includes lossless recordings for subjective audio quality testing. A version published in EBU Tech 3253 [30] was used. In terms of complexity, the resulting set includes 27 types of audio signals typically deployed in the context of audio broadcasting systems and web-casting applications. It can be divided into two principal categories, music and speech, at a ratio of 12 (music) and 15 (speech). Each recording is roughly 10–15 seconds long and sampled at 48 kHz.

For further work with the database, it was necessary to evaluate the audio quality of the selected signals, characterizing the typical content of the current audio broadcasting systems and web-casting applications, degraded by the codecs currently deployed by the corresponding systems and applications. In this case, the default codec settings were deployed besides the bit rate, which was manipulated according to the values presented in Tab. 1. We have chosen POLQA Music [7] model to assess a quality perceived by the end user when it comes to the degradations caused by the audio codecs, as this model, according to [7], was fairly reliable in this regard. It should be noted here that the POLQA Music V2 model was used in this case, see [7] for more information. The total number of combinations resulting from this process was 1080. These 1080 samples, together with the corresponding MOS values predicted by the POLQA Music model, represented a broadcast sub-database and were later used to build a web-cast sub-database covering the additional degradations, i.e., initial delay and stalling. As these two types of degradations are different in nature, i.e., frequency and time domain, in this case, we can apply the additivity concept that comes from the

E-model [17]. Since POLQA Music was not trained for degradations induced by the initial delay and stalling, we used the following equations published in [27], focusing on the effect of initial delay and stalling in the context of the audio transmission, to cover the impact of these two parameters on the quality experienced by the end user in the context of the model design:

$$f(T_{\text{Init}}) = 5 - (-0.5853 \log(T_{\text{Init}} + 1.3561) + 5), \quad (1)$$

$$f(T_{\text{Stall}}) = 5 - (3.3522e^{-0.6973 \cdot T_{\text{Stall}}} + 1.5059), \quad (2)$$

where T_{Init} represents the duration of the initial delay and T_{Stall} represents the duration of the stalling. In order to reflect real-world conditions, we have used the real-world measurements presented in [25] and [26] to select the values of initial delay and stalling to be used in the design process of the proposed models. The values, together with the corresponding quality degradations incurred by the initial delay and stalling, are listed in Tab. 2.

Tab. 2: The initial delay and stalling values used in the model design.

Type of audio signal degradation	Duration of degradation (s)	The value by which the original MOS is reduced
Initial delay	2	0.30777
	8	0.56838
	24	0.82188
Stalling	1	1.82494
	3	3.08026
	6	3.44301

To create the web-cast sub-database, we have extended the broadcast sub-dataset with the MOS values that cover the impact of initial delay and stalling on the quality perceived by the end user. It is necessary to note here that a combined impact of the initial delay and stalling was also considered while creating the web-cast sub-database. So, in total, we applied 15 cases, i.e., three initial delay cases (see Tab. 2 for more detail), three stalling cases (see Tab. 2 for more detail) and their nine combinations, to each of the previous 1,080 MOS samples representing the broadcast sub-database. Together, we obtained 17,280 MOS values defining the web-casting dataset. It is worth noting here that the database is available upon request to allow reproducibility of this research as well as support new research/development activities in this context.

4. Proposed Technique

In this work, we used a machine learning method, namely Random Forest (RF) [21] for regression problem to predict audio quality for broadcast systems and web-cast applications. The random forest can be applied to all kinds of regression applications. Random

regression forests are based on regression trees that lead to a minimization of variance within the node in the splitting process [21]. A decision tree is the building block of a random forest and is an intuitive model [21]. Instead of learning a simple problem, we will use a real-world dataset split into a training and testing set. In a random forest, a large number of trees is built from a random selection of a small number of variables and a random selection from the observations. The forest prediction is the average prediction from individual trees. Each individual tree predicts a step function. Even the average number of trees can approach almost any functional form and can automatically account for interactions between regressors. The most widely used variable importance metric for regression forests is permutation-based MSE (Mean Squared Error) reduction [23].

4.1. Training and Prediction Details

The sub-databases were divided into two parts at a ratio of 80:20, i.e., the training and testing data. It is worth noting here that when it comes to machine learning approaches, this ratio is often implemented. We pseudo-randomly selected the test cases/samples, attempting to cover all the types of signals, all the codecs and their bit rates at the same ratio. The testing samples were naturally not involved in the training database. The 240 MOS values for broadcast and 3,456 MOS values for web-cast were used for the testing phase to verify the performance of the designed parametric estimation models. The MOS values provided by the designed models were compared to the ground true MOS values. The efficiency of the constructed 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.

5. Results

To solve the regression problem, we used the existing random forest algorithm included in the Scikit-Learn Python library, namely "sklearn.ensemble.RandomForestRegressor" with the following parameters: the number of trees was set to 100, the random state was set to 50, maximum depth of the tree was set to default, i.e., "None", a minimum number of samples to split was set to 5, and a minimum number of samples required to be at a leaf node was set to 4. We ran 15 simulations for each sub-database with different initial seeds. The best results were noted and are going to be presented in this section. We compared the MOS values provided by the parametric prediction models with the true

MOS values of the test sets. The scatter plot of the true MOS values versus the predicted MOS values of the test samples obtained from the proposed models is shown in Fig. 3 for the broadcast sub-database and in Fig. 4 for the web-cast sub-database. Table 3 shows the best outcomes obtained from all the simulations in terms of the above-mentioned performance measures. The results show that the Random Forest method is relatively successful in both cases, i.e., the broadcast and web-cast conditions. To be more precise, for the web-cast conditions, we reached the PCC of 0.9854 and the RMSE of 0.2192; for the broadcast conditions, it is a bit worse with the PCC of 0.9411 and the RMSE of 0.1951, respectively.

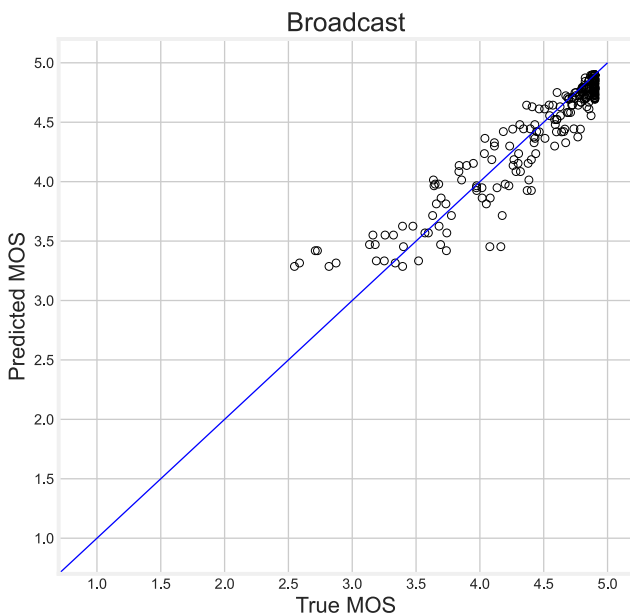


Fig. 3: Scatter plot of the true MOS values against the predicted MOS obtained by the Random Forest approach for the broadcast conditions.

Tab. 3: Best results obtained by the Random Forest approach.

Sub-database	Broadcast	Web-cast
PCC	0.9411	0.9854
RMSE	0.1951	0.2192

We also measured a computational load for both model types. The estimation computational load for the broadcast and web-cast sub-database is 0.0655 s and 0.2855 s respectively. All the simulations were carried out on a 64-bit quad-core processor based on the Kaby Lake H Architecture, Intel i7-7700HQ 2.8 GHz.

As we can clearly see from Fig. 3 and Fig. 4, there are some outliers for both types of environments. In the broadcast scenario, these outliers represent mostly speech signals and are covered by MP2, Opus, AAC-LC and HE-AAC (v.2) codecs at the lower bit rates. On the other hand, in the web-cast scenario,

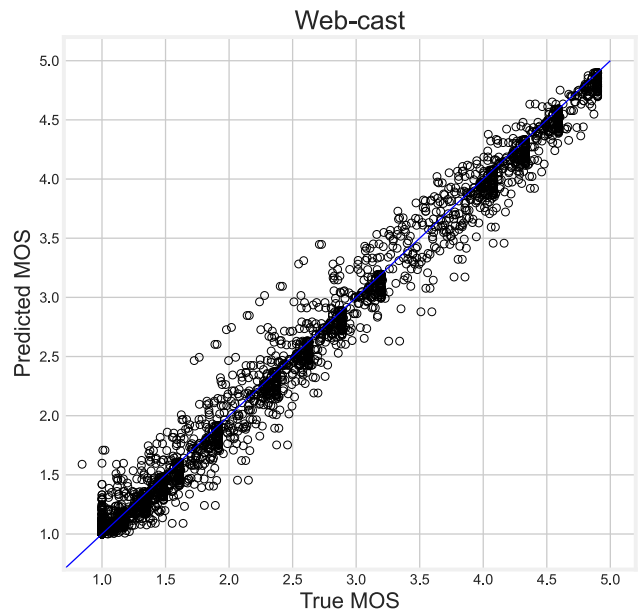


Fig. 4: Scatter plot of the true MOS values against the predicted MOS obtained by the Random Forest approach for the web-cast conditions.

there is an even representation of individual types of signals. Codecs that represent the largest share of outliers are MP2, AAC-LC, Ogg Vorbis, Opus and HE-AACv2, again for the lower bit rates.

6. Conclusion

In this work, we designed the non-intrusive parametric audio quality estimation models for broadcasting and web-casting scenario respectively, which are based on the Random Forest approach. The proposed models use the set of the audio features to estimate the overall audio quality for broadcasting systems and web-casting applications. When comparing the estimates provided by the proposed parametric models with the actual MOS values, we can conclude that the proposed approach estimates the perceived audio quality with a rather good accuracy. The developed parametric models can be implemented in monitoring systems, so that the quality of sound transmission can be calculated simultaneously on a large amount of connections. The future work will focus on the design of non-intrusive parametric audio quality estimation models based on other types of machine learning approaches, e.g. shallow and deep neural networks.

References

[1] ETSI Standard EN 300 401. *Radio Broadcasting Systems; Digital Audio Broadcasting (DAB)*

- to mobile, portable and fixed receivers. Valbonne: ETSI, 2006.
- [2] FISCHER, W. Digital Audio Broadcasting - DAB/DAB+. *Digital Video and Audio Broadcasting Technology*. 4th edition. Cham: Springer, 2020, pp. 575–622. ISBN 978-3-030-32184-0.
- [3] ETSI Standard ES 201 980. *Digital Radio Mondiale (DRM); System Specification*. Valbonne: ETSI, 2012.
- [4] HOFMANN, F., C. HANSEN and W. SCHAFER. *Digital Radio Mondiale (DRM) digital sound broadcasting in the AM bands*. *IEEE Transactions on Broadcasting*. 2003, vol. 49, iss. 3, pp. 319–328. ISSN 1557-9611. DOI: 10.1109/TBC.2003.818718.
- [5] SEUFERT, M., S. EGGER, M. SLANINA, T. ZINNER, T. HOSSFELD and P. TRAN-GIA. A Survey on Quality of Experience of HTTP Adaptive Streaming. *IEEE Communications Surveys & Tutorials*. 2015, vol. 17, iss. 1, pp. 469–492. ISSN 1553-877X. DOI: 10.1109/COMST.2014.2360940.
- [6] HOSSFELD, T., M. SEUFERT, M. HIRTH, T. ZINNER, P. TRAN-GIA and R. SCHATZ. Quantification of YouTube QoE via Crowdsourcing. In: *2011 IEEE International Symposium on Multimedia*. Dana Point: IEEE, 2011, pp. 494–499. ISBN 987-1-4577-2015-4. DOI: 10.1109/ISM.2011.87.
- [7] 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.
- [8] STOCKHAMMER, T. Dynamic adaptive streaming over HTTP –: standards and design principles. In: *MMSys '11: Proceedings of the second annual ACM conference on Multimedia systems*. New York: ACM, 2011, pp. 133–144. ISBN 978-1-4503-0518-1. DOI: 10.1145/1943552.1943572.
- [9] BERG, J., C. BUSTAD, L. JONSSON and L. MOSSBERG. 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.
- [10] CAMPBELL, D., E. JONES and M. GLAVIN. Audio quality assessment techniques—A review, and recent developments. *Signal Processing*. 2009, vol. 89, iss. 8, pp. 1489–1500. ISSN 0165-1684. DOI: 10.1016/j.sigpro.2009.02.015.
- [11] BS.1116. *Methods for the Subjective Assessment of Small Impairments in Audio Systems*. Geneva: ITU-R Recommendation, 2015.
- [12] BS.1534. *Method for the Subjective Assessment of Intermediate Quality Levels of Coding Systems*. Geneva: ITU-R Recommendation, 2015.
- [13] BS.1387. *Method for objective measurements of perceived audio quality*. Geneva: ITU-R Recommendation, 2001.
- [14] HUBER, R. and B. KOLLMEIER. PEMO-Q—A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception. *IEEE Transactions on Audio, Speech, and Language Processing*. 2006, vol. 14, iss. 6, pp. 1902–1911. ISSN 1558-7924. DOI: 10.1109/TASL.2006.883259.
- [15] HINES, A., E. GILLEN, D. KELLY, J. SKOGLUND, A. KOKARAM and N. HARTE. ViSQOLAudio: An objective audio quality metric for low bitrate codecs. *The Journal of the Acoustical Society of America*. 2015, vol. 137, iss. 6, pp. 449–455. ISSN 1520-8524. DOI: 10.1121/1.4921674.
- [16] RIX, A. W., J. G. BEERENDS, D.-S. KIM, P. 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.
- [17] ITU-T G.107. *The E-model: a computational model for use in transmission planning*. Geneva: ITU-T Recommendation, 2015.
- [18] GRAUBNER, M., P. S. MOGRE and T. LORENZEN. QoE Assessment for Broadcast Audio Contribution over IP (ACIP). In: *AES 38th International Conference on Sound Quality Evaluation*. Pitea: AES, 2010, pp. 1–10.
- [19] GRAUBNER, M., P. S. MOGRE, R. STEINMETZ and T. LORENZEN. A new QoE model and evaluation method for broadcast audio contribution over IP. In: *NOSSDAV '10: Proceedings of the 20th international workshop on Network and operating systems support for digital audio and video*. Amsterdam: Association for Computing Machinery, 2010, pp. 57–62. ISBN 978-1-4503-0043-8. DOI: 10.1145/1806565.1806581.
- [20] ITU-T P.1201. *Parametric non-intrusive assessment of audiovisual media streaming quality*. Geneva: ITU-T Recommendation, 2012.

- [21] BREIMAN, L. Random Forests. *Machine Learning*. 2001, vol. 45, iss. 1, pp. 5–32. ISSN 1573-0565. DOI: 10.1023/a:1010933404324.
- [22] LAVNER, Y. and D. RUINSKIY. A Decision-Tree-Based Algorithm for Speech/Music Classification and Segmentation. *EURASIP Journal on Audio, Speech and Music Processing*. 2009, vol. 2009, iss. 1, pp. 1–14. ISSN 1687-4714. DOI: 10.1155/2009/239892.
- [23] GROMPING, U. Variable importance in regression models. *WIREs Computational Statistics*. 2015, vol. 7, iss. 2, pp. 137–152. ISSN 1939-0068. DOI: 10.1002/wics.1346.
- [24] 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.
- [25] SCHWIND, A., L. HABERZETTL, F. WAMSER and T. HOSSFELD. QoE Analysis of Spotify Audio Streaming and App Browsing. In: *Internet-QoE'19: Proceedings of the 4th Internet-QoE Workshop on QoE-based Analysis and Management of Data Communication Networks*. Los Cabos: Association for Computing Machinery, 2019, pp. 25–30. ISBN 978-1-4503-6927-5. DOI: 10.1145/3349611.3355546.
- [26] 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.
- [27] 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.
- [28] YAMAGISHI, K. and T. HAYASHI. Parametric Quality-Estimation Model for Adaptive-Bitrate-Streaming Services. *IEEE Transactions on Multimedia*. 2017, vol. 19, iss. 7, pp. 1545–1557. ISSN 1941-0077. DOI: 10.1109/TMM.2017.2669859.
- [29] SEUFERT, M., P. CASAS, N. WEHNER, L. GANG and K. LI. Stream-based Machine Learning for Real-time QoE Analysis of Encrypted Video Streaming Traffic. In: *2019 22nd Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN)*. Paris: IEEE, 2019, pp. 76–81. ISBN 978-1-5386-8336-1. DOI: 10.1109/ICIN.2019.8685901.
- [30] Sound Quality Assessment Material recordings for subjective tests EBU – TECH 3253. In: *EBU Technology & Innovation* [online]. 2008. Available at: <https://tech.ebu.ch/docs/tech/tech3253.pdf>.

About Authors

Martin JAKUBIK was born in Ruzomberok, Slovak Republic. He received his M.Sc. in Telecommunications from the 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 prediction using machine learning methods.

Peter POCTA was born in 1981. He received his M.Sc. 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 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 refereed papers in international journals and conferences including Acta Acustica united with Acustica, Applied Acoustics (Elsevier), Speech Communication (Elsevier), IEEE Transactions on Broadcasting, MESAQIN and 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.