Advances in Computer Science Research, volume 89
2019 International Conference on Wireless Communication, Network and Multimedia Engineering (WCNME 2019)

Objective Quality Measurement in Multi-channel Audio Systems by Multivariate Adaptive Regression Splines Model

Huaping Liu and Yong Fang*
Shanghai Institute for Advanced Communication and Data Science, Key laboratory of Specialty Fiber Optics and Optical Access Networks, Joint International Research Laboratory of Specialty Fiber Optics and Advanced Communication, Shanghai University, Shanghai, 200444, P. R. China
*Corresponding author

***Abstract*—Objective quality assessment methods have been used widely for evaluation of audio systems. This article introduces a new method to show the relationship between the input parameters and the prediction targets of the multivariate adaptive regression splines (MARS) model. In this proposed method, the relatively frequency of each input variables selected by the MARS model as useful predictors is calculated. The MARS model is trained and tested by the ITU DB4 database which is generated by ITU-R WP6C in order to evaluate the high-quality multichannel audio coding approaches. The under-test input parameters consist of one processed monaural feature and several unprocessed binaural features, the prediction target is the subjective quality score of test items. The proposed method indicates that the binaural features are also important, although the contribution made by them are relatively low.**

Keywords—multivariate adaptive regression spline; subjective quality; monaural feature; binaural feature

I. Introduction

The MARS[1] model is a non-parametric regression model popularized by Friedman for modeling of high dimensional data. It allows the researchers to model an outcome as a function of one or more input variables. These input parameters are also called predictors. The MARS model can be seen as an extension of linear regression models and takes the form of spline basis functions (BFs). Its model building procedure contains two passes: a forward pass and a backward pass. In the forward pass, an approximation model is built from the input series. This model is usually over fitted, thus it should be generalized in the backward pass by deleting the less important basic functions. An explicit expression together with the model parameters can be obtained after the backward pass. In other words, only useful input variables will be selected in the model building procedure.

The proposed objective quality[2] measurement in improved method implements the MARS model to map the selected objective quality degradation variables to the normalized subjective scores. In the experimental part, the MARS model is trained firstly and the training procedure is repeated 500 time. The MARS model with best prediction accuracy is chosen for the test procedure. The training and testing data belong to the same database.

The input parameters of the MARS model are one monaural feature and three binaural features. The monaural feature is already mapped from the 5 timbral parameters in the advanced perceptual evaluation and audio quality (PEAQ) model by a trained multi-layer network (NN). It has the capability to represent the timbral quality degradation of the testes signals. Other inputs are the interaural level difference degradation (ILDD), the interaural time difference degradation (ITDD) and the interaural cross-correlation degradation (IACCD)[3], respectively. They are calculated from the corresponding differences between original and test signals.

There are two methods for evaluating the contribution made by different input features to build the MARS model. Both of them show that the most contribution is given by the monaural feature. Although some of the binaural features are chosen as useful predictors, their contributions are still low. It indicates that these methods are not suitable for the cases where one parameter has very high contribution to the model compared to others. To address this problem, a new method introduces a parameter “ratio” to record the relative frequency of the MARS model taking monaural or binaural features as useful inputs.

This paper is organized as follows: Section II introduces the multivariate adaptive regression splines model, Section III describes the optimization method and constructs the proposed evaluation method, Section IV evaluates the optimization performance. Section V concludes this paper.

II. The multivariate adaptive regression spline model

The regression splies model maps the multiple features into an outcome which should have the capability to predict the subjective scores from the listening tests. As shown in Eq. (1), the predicted relation f(x) consists of several basis functions, each of which is multiplied with a constant weight cm.

$$f(x) = \sum_{m=1}^{M} c_m B_m(x) \qquad (1)$$

For piece-wise linear regression, the basis function Bm(x) can be a constant, a hinge function or the product of two or more hinge functions. Hinge function takes the form as max(0;;

x-c) or max(0; c-x) where c is called as knot. Figure 1 shows a mirrored pair of hinge functions with a knot at 1.5. For piece-wise cubic regression, the hinge function is replaced by the cubic splines function.

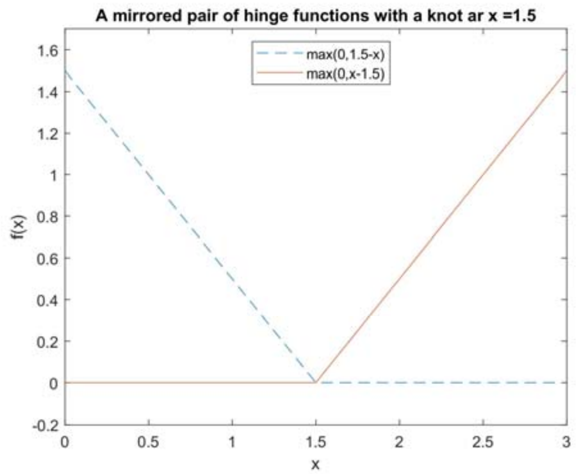

FIGURE I. AN EXAMPLE OF MIRRORED HINGE FUNCTIONS

The term named as generalized cross validation (GCV) error is a vital criterion for choosing the model with satisfying generalization ability in the backward pass. The method for calculating the GCV error is shown in Eq.(2). The numerator represents the mean square error (MSE) of prediction. Besides, the denominator corresponds to the model complexity where M is the number of basic functions, d is a penalty for each basis function and (M-1)/2 is the number of knots. Thus, there is a trade-off between the fitting performance and model complexity.

$$\mathrm{GCV} = \frac{\frac{1}{N}\sum_{i=1}^{N}[y_i - f_i(x)]^2}{1-\frac{M+d(M-1)/2}{N}} \qquad (2)$$

III. THE PROPOSED EVALUATION METHOD

As shortly described in the previous section, the newly designed method is used to tell whether the input variables are important or not with the help of factor “ratio”. In this section, the basic structure of the experiment subject will be introduced first. Then, the way to calculate “ratio” will be presented.

A. The Introduction to the Audio Objective Quality Measurement

The objective quality measurement is carried out through algorithms that process the signals under test using a perceptual model of the human auditory system. It provides a quality score that should predict the results of listening tests. The listening test is a gold standard for audio quality evaluation, in which the experienced listeners are required to rate the specific stimuli or detect the quality degradation of test signals compared to the reference signal. However, such kind of subjective quality measurement procedure is time-consuming and lacks flexibility. Therefore, objective quality assessment has received increasing attention in recent years.

Perceptual evaluation and audio quality (PEAQ)[4] is a widely used evaluation system, which measures the perceived audio quality by taking the characteristics of human auditory system into account, according to ITU-R BS.1387-1. The basic structure of the PEAQ system is illustrated in Figure 2.

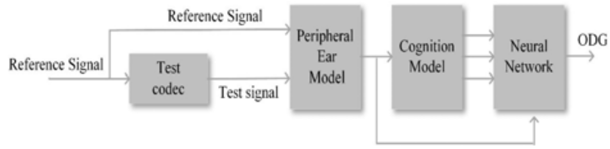

FIGURE II. BASIC STRUCTURE OF THE PEAQ SYSTEM

The final output of PEAQ is called objective difference grade (ODG) mapped from the model output variables (MOVs) generated by the cognition model. These ODGs should mimic the behavior of the subjective scores and therefore estimate the perceived quality degradation of test signals. Unfortunately, PEAQ was designed for mono and stereo audio signals and this purely technical measure is not able to describe the objective quality degradation in spaciousness. Therefore, some modifications need to be made on the basis of the PEAQ system. In the new objective quality measurement, an extended model is added on the basic structure of PEAQ, which extracts three binaural features: ILDD, ITDD and IACCD to represent the spaciousness degradation. The timbral degradation is still calculated by the traditional PEAQ structure and marked by one monaural feature called monaural ODG[6].

The MARS model is used to map the monaural ODG and three binaural features into an estimated ODG. The main reason for replacing neural network with the MARS model is lower complexity.

B. The Parameter Importance Evaluation

The existing two evaluation methods are described by Friedmann in [1]. The first one is denoted as ANOVA decomposition[1], which is used for interpreting the influence of these inputs to the outputs. The second method uses a term called delGCV to describe the importance of relative variables[5]. The maximum value of delGCV is 100 and if delGCV was equal to 0, it means that the corresponding input variable is not used by the MARS model. In this measure, the best predictor gets a delGCV of 100 and the less important predictors obtain smaller delGCVs. These two methods focus on quantifying the contribution given by different inputs. It is apparent that the monaural feature is the most valuable input for predicting the corresponding subjective scores. However, the negligible contributions from other binaural features do not mean that they are not related to the quality degradation of test signals.

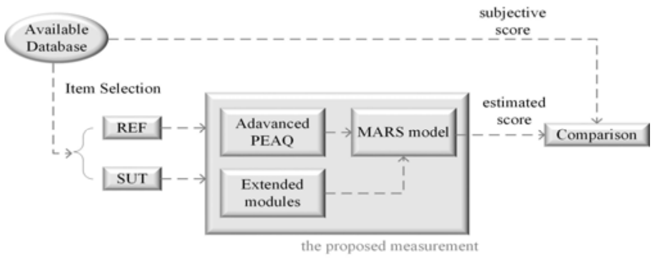

FIGURE III. THE CHART FOR THE TRAINING PROCEDURE OF THE MARS MODEL

IV. EVALUATION OF EXPERIMENTAL RESULTS

In order to obtain a reliable MARS model, data from the available database are partitioned in two different groups, 50% of the data will be used for training the MARS model and the rest will be used for verification. As shown in Figure 3, the subjective data corresponding to the items of the training group will be fed to the regression model, which in turn will map input features calculated on these items to said subjective scores. The outputs of trained MARS models will be called estimated scores and contemplates binaural as well as monaural distortions. The estimated scores are compared to the truly subjective scores by using Eq.(3).

$$R = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y} \qquad (3)$$

This training procedure will be repeated 500 times, thus there will be 500 MARS models. As introduced previously, only the useful input parameters will be appeared in the final expression of the MARS model after the forward and backward pass. The times of each input variables chosen as useful parameters will be accumulated and then divided by the total repetition number. The value is called “ratio”. Then, the binaural features are replaced by noise-like vectors, such as zero or random values within [0;1], in order to reflect the importance of them. Comparing the ratio values of three binaural features with those of noise-like inputs can help us understand their importance.

The available database is extracted from the ITU DB4 database which is generated by ITU-R WP6C[8]. There are 40 multichannel audio items that are roughly classified into three parts: music, movie and ambience signals. These original items were processed in 4 different conditions, which are listed in Table.1. Thus, there are 160 sound excerpts in total[9].

TABLE I. CONDITION LIST OF THE AVAILABLE DATABASE

Condition	Bitrate
Dolby AC3	320kbps
MPEG Surround(Layer II)	256kbps
AAC LD	480kbps
MPEG Surround(AAC-LC core)	192kbps

The subjective assessment of this database is performed by the BS.1116 test[7]. The subjective score is calculated by subtracting the grade of test and reference signal. The objective features are generated by the objective quality measurement. The proposed evaluation method is tested by the objective features and subjective scores of this database[10][11].

The results are shown in Table.2, the ratio of objective features are much higher than noise or zero. It meets the expectation before performing this experiment. Those binaural features: ILDD, ITDD and IACCD are also related to the perceivable quality degradation of test signals. The prediction result of the selected MARS model from the 500 repetitions is close to 94%, which is a pretty high value. Therefore, this method can be used to show the relationship between test input variables and prediction target.

TABLE II. CONDITION LIST OF THE AVAILABLE DATABASE

Input variables	Ratio
Monaural Feature	100.0%
ILDD	93.0%
ITDD	91.6%
IACCD	85.2%
Noise	52.4%
Zero	0.0%

V. CONCLUSION

In this paper, an objective method is introduced that can be used to predict perceived quality in multi-channel audio compression coding systems based on MARS. The method takes into account degradations in both spatial quality and timbral quality, extending previous approaches by incorporating a binaural-hearing model from which interaural feature are computed. The MARS model is used to map the monaural ODG and three binaural features into an estimated ODG. The main reason for replacing neural network with the MARS model is lower complexity.

ACKNOWLEDGMENT

The work was supported by the Key Support Projects of Shanghai Science and Technology Committee (16010500100).

REFERENCES

[1] J.H. Friedman. Multivariate adaptive regression splines. The Annals of Statistics, 1991.

[2] J.-H. Seo, S.B. Chon, K.-M. Sung, and I. Choi. Perceptual objective quality evaluation method for high quality multichannel audio codec, J. Audio Eng. Soc, 2013.

[3] J.-H. Flebner, R.Huber, and S.Ewert. Assessment and prediction of binaural aspects of audio quality, Journal of the Audio Engineering Society, 2017.

[4] M.Schafer, M. Bahram, and P. Vary. An extension of the PEAQ measure by a binaural hearing mode. IEEE International Conference on Acoustics, Speech and Signal Processing 2013.

[5] G. Jekabsons, ARESLab: Adaptive regression splines toolbox for Matlab/Octave, 2016.

[6] F. Baumgarte and C. Faller, Binaural cue coding-part I: Psychoacoustic fundamentals and design principles. Speech and Audio Processing, IEEE Transactions on, 2003.

[7] I. JTC1/SC29/WG11, International organization for standardization organization international demoralization ISO/IEC JTC1/SC29/WG11 coding of moving pictures and audio, Geneva, Switzerland, 2011.

[8] I.-R.R.BS.1534-1, Method for the subjective assessment of intermediate quality levels of coding systems. International Telecommunication Union, Geneva, Switzerland, 2003.

[9] I.-R.W.D.C.R.A. 13, Progress report toward revision of recommendation ITU-R BS.1387-1 method for objective measurements of perceived

audio quality. International Telecommunication Union, Geneva, Switzerland, 2010.

[10] J. Culling, Subcomponent cues in binaural unmasking. The Journal of the Acoustical Society of America, 2011.

[11] I. Choi, B.G.Shinn-Cunningham, S.B.Chon and K.-M. Sung. Objective measurement of perceived auditory quality in multi-channel audio compression coding systems. J. Audio Eng. Soc., vol56, pp. 3~17, 2008.