# THE LASSO AND ELASTIC-NET ALGORITHMS FOR PREDICTIVE SOUND QUALITY MODELS USING LARGE POOL OF PREDICTIVE METRICS AND FACTORS

**Abdelghani Benghanem** *[1], **Philippe-Aubert Gauthier** †[1,2], **Lévy Leblanc** ‡[1], **and Alain Berry** §[1,2]
[1]Groupe d'Acoustique de l'Université de Sherbrooke, Sherbrooke, Québec, Canada
[2]Centre for Interdisciplinary Research in Music, Media and Technology, McGill University, Montréal, Canada

## 1 Introduction

Part of the perceived quality is communicated by sound quality (SQ). SQ studies present a twofold challenge [1]. The first is that they rely on listening tests. However, listening tests are considered time-consuming. The second challenge relates to the development of predictive SQ models. Often, these prediction models are derived using linear regression on a limited set of predictive metrics. However, nowadays, many metrics are available and there is no computational burden that should limit the number of potential metrics. An issue is that regression using more metrics than observations cannot lead to meaningful predictive models since all the metrics are selected. In this paper, different algorithms are compared to construct a SQ predictive model that does not suffer from these limitations. These algorithms achieve an automatic selection of few metrics from a large pool. The lasso [2] and elastic-net [3] are tested for the prediction of listening tests results of consumer product.

## 2 Listening tests: Method

Listening tests were conducted with 40 participants according to a forced-choice pairwise comparison method [1]. Tests were conducted using calibrated and equalized binaural reproduction over headphones. The test was based on a full-fractional design of experiments (DOE) with four factors (two levels). The sound samples were based on a 5-second binaural sound that has been transformed into 16 different sounds according to the DOE using parametric time-varying equalizers. Levels were corresponding to increased or decreased content corresponding to factors. The participants controlled the playback and their selection with a user interface. The pairwise comparison tests included 120 pairs that were compared with random permutation each time.

## 3 Lasso and elastic-net algorithms

The aim of SQ studies is the derivation of predictive SQ models based on results of listening tests (herein the merit values $v_i$ of sound sample $i$) and potential predictive factors or metrics derived from the actual sound samples or from DOE factors. A common form of predictive SQ model is expressed as follow for each of the $I$ merit values

$$v_i = \sum_{m=1}^{M} F_{im} b_m + b + e_i \qquad (1)$$

with merit values $v_i$ obtained from listening test statistics, matrix of $M$ factors and/or metrics $F_{im}$, $M$ linear regression coefficients $b_m$, $b$ is the intercept and $e_i$ is the prediction error. Note that the matrix of factors and predictors $F_{im}$ can not only include factors and metrics, but interactions of factors and non-linear functions of metrics (squared, log, etc.). In this case, although the model of expression Eq. (1) is linear, the actual SQ model is not. However, Eq. (1) can still be solved with linear algorithm: using linear regression. For the case of *wide data* with much less observations than predictive metrics (i.e. $I << M$), classical linear regression algorithms will typically achieve over-fitting and all the coefficients $b_m$ will be non-zero. Hence limiting the interpretable meaning of the resulting model for any engineering uses or design guidelines. The aim of the lasso and elastic-net algorithm is to circumvent this issue by rewriting the problem as a composite cost function:

$$J_{\lambda,\alpha} = \frac{1}{2I}\left(\sum_{i=1}^{I}\left(v_i - b - \sum_{m=1}^{M} b_m F_{im}\right)\right)^2 + \lambda \sum_{m=1}^{M}\left(\frac{(1-\alpha)}{2} b_m^2 + \alpha |b_m|\right) \qquad (2)$$

The first right-hand side term corresponds to the quadratic sum of the predictor errors and the second right-hand side term is a regularization term with regularization amount $\lambda > 0$. The regularization combines 2-norm regularization $b_m^2$ and 1-norm regularization $|b_m|$ based on the elastic-net parameter $0 \leq \alpha \leq 1$. For the lasso algorithm, $\alpha = 1$, only the 1-norm regularization is included. This induces solution sparsity, i.e. few coefficients $b_m$ will be non-zero and selected in the model [2]. The sparsity is controlled by the regularization amount $\lambda$. The elastic-net algorithm [3] involves $0 < \alpha < 1$ and it will typically smooth out the selection towards the extreme case of 2-norm-only regularization with $\alpha = 0$ (this corresponds to Tikhonov regularization). The minimum of the composite cost-function cannot be found by analytical means since it is non-linear. Therefore, this is solved iteratively using coordinate-descent algorithm. $\alpha$ is a user-selected parameter. The penalization amount $\lambda$ must be set rigorously: The experimenter defines a limit on the maximum number of metrics that should be included. Next, the problem is solved for a wide range of $\lambda$ starting with the largest $\lambda$ so that all coefficients $b_m$ are zeros. Then the $\lambda$ iteration stops if: 1) the maximum number of non-zero coefficients $b_m$ is reached or 2) the minimum prediction error is obtained for $k$-fold cross-validation of the prediction. In this paper, the maximum number of predictive metrics is 10 and the cross-validation is 8-fold.

∗. Abdelghani.Benghanem@USherbrooke.ca
†. philippe_aubert_gauthier@hotmail.com
‡. Levy.Leblanc@USherbrooke.ca
§. Alain.Berry@USherbrooke.ca

## 4 Factors and metrics

Various predictors $F_{im}$ were used. The four DOE factors (namely $F1$, $F2$, $F3$, and $F4$) were included. Next, interactions of factors were also included ($F1F2$, $F1F3$, etc. for a total set of 14 predictive interaction factors). Spectral informations was gathered as predictive metrics: 1) third-octave-band levels, 2) specific loudness on Bark scale. Finally, other metrics and related time statistics (mean value, standard deviation, and slope) were included using the MIR Toolbox (Music Information Retrieval) [4]. All metrics and potential predictors were centered and normalized before the construction of the predictive SQ model. The pool of predictors included 141 metrics and factors.

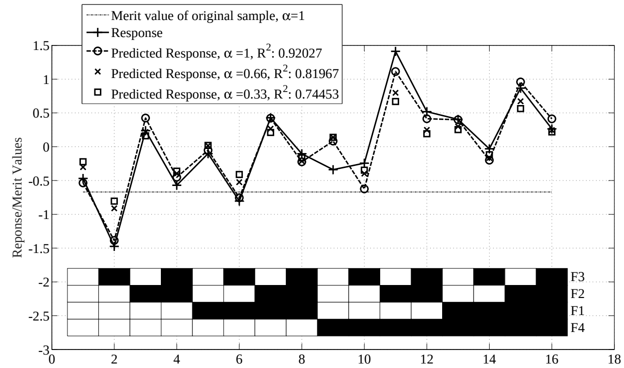
## 5 Post-processing and results

Once listening tests were completed, few post-processing steps were done: 1) participants clustering using a Gaussian mixture model in order to extract the most coherent cluster of participants, 2) one-way ANOVA was performed, the obtained $p$-value is 9.67e-31 which is very low, hence validating the test (i.e. the presented sound samples had an effect), and 3) the merit values $v_i$ were obtained based on the probability matrix and Gumbel distribution [1].

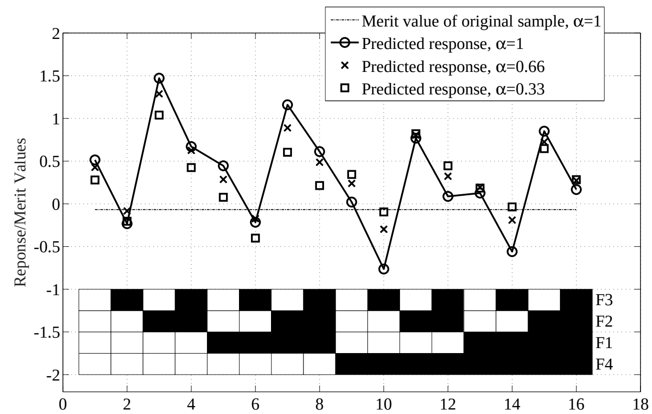
Merit values $v_i$ and predicted merit values are reported in Fig. 1(a) for the lasso and elastic-net algorithms. Excellent predictions of the merit values are observed. For the lasso, only 4 metrics were selected. For the elastic-net, 5 or 8 predictors were selected in the final SQ model for $\alpha = 0.66$ and $\alpha = 0.33$, respectively. The three algorithms predict very similar merit values, the major difference between the algorithms is not related to prediction accuracy but with the number of selected predictors. It is worth noting that for the lasso (which provides the most sparse predictor selection) none of the DOE factors were selected, i.e. signal processing and psychoacoustic metrics are judged as more appropriate for the predictive SQ model. An illustrative example for which the usefulness of the prediction model is shown in Fig. 1(b). The three obtained predictive SQ models for Fig. 1(a), as obtained for the described listening tests, were applied to another set of 16 sounds based on a different original sound (same kind of product, different model) that has been through modifications as reported for the DOE.

## 6 Conclusions

The lasso and elastic-net algorithms were compared for the prediction of listening tests results of consumer product for which 141 metrics and DOE factors were used as predictors. It was shown that the most promising algorithm is the lasso which is able to efficiently and strictly limit the number of metrics and provide a model that can be used in order to provide a set of understandable design guidelines. Typically, the elastic-net includes more predictors in the predictive SQ model.




(a) $v_i$ and predicted $v_i$ for the 16 sound samples used in listening tests.




(b) Predicted $v_i$ of the new 16 sound samples.

**Figure 1:** (a) Merit values and predicted merit values for the predictive SQ models using the lasso ($\alpha = 1$) and elastic-net ($\alpha = 0.33$ and $\alpha = 0.66$) for the 16 sound samples used in listening tests. (b) Predicted merit values of the new 16 sound samples for the predictive SQ models using the lasso ($\alpha = 1$) and elastic-net ($\alpha = 0.33$ and $\alpha = 0.66$) algorithms. Factors of the DOE are shown as black and white patches. Black and white patches mean high and low levels, respectively.

## Acknowledgments

This work was financially supported by the National Sciences and Engineering Research Council of Canada, BRP and CTA for the dSkibel research project.

## References

[1] N.C. Otto and G.H. Wakefield. A Subjective Evaluation and Analysis of Automotive Starter Sounds. *Noise Control Engineering Journal*, 41(3):377, 1993.

[2] R. Tibshirani. Regression Selection and Shrinkage via the Lasso. *Journal of the Royal Statistical Society*, 58:267–288, 1994.

[3] H. Zou and T. Hastie. Regularization and variable selection via the elastic-net. *Journal of the Royal Statistical Society*, 67:301–320, 2005.

[4] P. Lartillot O., Toiviainen. A Matlab Toolbox for Music Feature Extraction from Audio. *Proceedings for the 10th International Conference on Digital Audio Effects (DAFx-07), Bordeaux, France*, 2007.