

VARIABLE SELECTION ANALYSIS FOR DECISION TREE REGRESSION MODELS OF SOUNDSCAPES EMOTIONS

PACS: 43.60.Np

San Millán-Castillo, Roberto; Martino, Luca; Morgado, Eduardo.
Universidad Rey Juan Carlos – ETSIT – Departamento de Teoría de la Señal y
Comunicaciones, Camino del Molino, 5, Fuenlabrada (Comunidad de Madrid, España),
roberto.sanmillan@urjc.es

Palabras Clave: Variable selection, feature selection, decision tree regression, soundscapes emotion recognition.

ABSTRACT.

Signal regression models are useful tools for prediction, interpolation and smoothing. Countless variables are considered to characterize acoustic signals. However, models with many variables may end up in less interpretable solutions and with a high computational load. Thus, variable selection studies are central to improving models' performance. This article analyzes a well-known database. It employs non-linear soundscape emotion models, such as decision tree regression, considering two outputs (soundscape descriptors): arousal and valence. We carried out the models' performance and variable selection analysis. The results show that a reduced space of features (soundscape indicators) can provide parsimonious models with competitive performance.

RESUMEN.

Los modelos de regresión son herramientas útiles para predecir, interpolar y suavizar señales. Incontables variables se extraen para caracterizar una señal acústica. Sin embargo, los modelos con muchas variables pueden generar soluciones menos interpretables y alta carga computacional. Así, la selección de variables es relevante para mejorar modelos. Este artículo analiza una conocida base de datos. Considera modelado no-lineal de las emociones percibidas de paisajes sonoros, árboles de decisión, de dos salidas (descriptores): *arousal* y *valence*. Se evaluó el rendimiento de los modelos y la selección de variables. Los resultados muestran que pocas variables (indicadores) proporcionarían modelos sencillos con métricas competitivas.

1. INTRODUCTION

Variable selection is a central task in signal processing, statistics, and machine learning. Before dealing with variable selection, extracting variables (i.e., features) to characterize signals, is a

straightaway process that may lead to uncountable variables describing the original signal. Generally, the target in variable selection is to only include the relevant variables that improve the model performance. When using many variables, models are likely to show overfitting. Thus, variable selection presents some advantages: (a) few features lead to simple models; (b) simple models mean low computational load; (c) and so, faster algorithms for real-time applications.

When it comes to the application, soundscapes are becoming one of the most active topics in acoustics nowadays. Soundscapes broaden the classical environmental acoustics vision beyond the idea of 'noise'. Soundscapes provide a holistic approach including individuals' perceptions, context, and acoustic environments. Soundscapes' modelling may help forecast human responses to different acoustic circumstances with few resources.

Soundscapes-elicited emotions play a relevant role in some applications such as urban planning, noise monitoring, sound design in films and digital games [1], or sonification [2]. Thus, soundscape emotion recognition (SER) is a relatively new sub-field of research with promising benefits. Following Russell's circumplex affect model, SER can be sufficiently modelled with two relevant factors: *arousal* and *valence*, which represent the eventfulness and the pleasantness ratio of an acoustic environment, respectively [3], [4]. An extensive range of soundscapes descriptors (i.e., outputs) and soundscapes indicators (i.e., variables/features) have been researched up to now. Non-linear models seem to provide better performance than linear models, which are preferred because they are simpler to develop [5]. Most of these studies do not share a comparison framework. This paper employs the Emo-soundscapes database (EMO) [6], which is being a reference in SER studies recently [2], [7]–[9].

This work aims to analyze the variables' importance of a model for SER, which is based on a non-linear approach such as Decision Tree Regression (DTR) model, which have been already employed in Acoustics [2], [10]. From a simplicity and interpretability point of view, DTR may be competitive with linear regression. Hence, the preliminary results and contributions of this study are the following:

- The evaluation of the performance of DTR models to EMO as an alternative to linear regression, random forest, support vector machines and artificial neural network strategies.
- The selection of variables of the model with an *embedded* approach, supported by their relevance by the *Gini importance* criterion.

The remainder of the paper is organized as follows. *Section 2.1* comprises a description of the database that was employed in this study. *Section 2.2* presents some background on DTR. *Section 2.3* describes the employed framework for variable selection. After that, *Section 3* shows the results applied to the database. Finally, *Section 4* draws some conclusions from the previous analysis.

2. MATERIALS & METHODS

2.1. Database

This study works with EMO, which is likely to be the largest publicly available database of soundscapes with annotations of emotion labels currently [10]. EMO consists of more than 1200 audio files under a Creative Commons license. The EMO's files are classified according to Schafer's taxonomy [11]. A crowd-sourcing procedure provides the perceived soundscape emotions, based on Russell's affect representation, by 1182 trusted annotators with adequate inter-subject reliability.

EMO provides up to 122 normalized variables/features that are extracted from every audio file, with a 50% overlapping Hanning window (23 ms wide). Among the range of variables, there are *Psychoacoustic features*, such as Loudness and MFCCs; *Time-domain features*, such as Energy and Entropy; and *Frequency-domain features*, such as Pitch and Centroid.

2.2. Decision Tree Regression

DTR is a non-linear and non-parametric supervised learning method that predicts the value of the selected output with simple decision rules, which are inferred from the input variables in the training data. DTR may present visible and easy interpretation of results using a tree structure, and low resources on data preparation. On the other hand, DTR may produce overfitting when decision rules become overcomplicated, and sensitive to small data changes and data imbalance. This study uses the CART algorithm [12], which works as follows: the input variable space is divided in M overlapping regions $R_m = \{1 \dots M\}$. For every observation, $n_{samples}$, that falls into R_m , CART predicts SER, let $y_{o,i}$ be the value of arousal or valence, as the mean of the output values of the training data in R_m , let $\widehat{y}_{o,i}$ be the predicted value of arousal or valence. DTR aims to search for the R_m that minimizes the Mean Square Error (MSE), which grows the tree according to (1). To this end, firstly, it finds the input variable V_m and cut-point p such, that the splitting of the variable space into regions $\{V | V_m < p\}$ and $\{V | V_m \geq p\}$ results in the maximum possible reduction in MSE (R_m). Next, the process repeats but splits the two previously identified regions. The process is an iterative greedy algorithm that remains until a stop is reached.

$$MSE(R_m) = \frac{1}{n_{samples}} \sum_{m=1}^M \sum_{i \in R_m} (y_{o,i} - \widehat{y}_{o,i})^2, \quad (1)$$

DTR deals with a range of hyperparameters for tuning the models' predictions. For the sake of simplicity, this work only varies one of the most relevant hyperparameters: the maximum number of splits for a sample (MaxDepth). The remainder of the DTR setup is as follows: (a) The best split considers all the input variables and (b) the minimum number of required samples to split a node is two. Besides, (c) leaf nodes are unlimited and are correct with only one sample.

2.3. Variable Selection Framework

The reduction of variables for a model can be performed by variable transformation or variable selection methods. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are examples of variable transformations that provide new variable sets from the original ones and their combinations. This study works with variable selection methods that consider subsets of the original variable space without any transformation through a classical procedure: (a) variable subset selection; (b) variable subset evaluation; (c) stop criterion setting; and finally, (c) the assessment of the results. The target is to maximize the relevance and minimize the redundancy of a variable set [13]. Variable selection methods are usually classified into *filters* (e.g., Correlation, Relief), *wrappers* (e.g., Naïve Bayes + Regression), and *embedded* (e.g., DTR, LASSO) methods, assuming variable independence [13]. Research in this field is fruitful but there is no general solution. Many authors agree methods may suit research problems depending on their features[14].

This paper presents a variable selection framework based on an *embedded* method, which is supported by DTR and a heuristic filter method. Firstly, the model's fitting by DTR provides all variables' importance/relevance. After that, variables are sorted according to their importance.

Then, the most relevant variables are selected to build up DTR models with a reduction of variables heuristically. This study employed the Gini Index or Gini Importance (GI) as the variable importance criterion, which calculates the times a variable is used to split a tree node (weighted by the number of samples of the node).

In the first stage, DTR models consider the 122 variables included in EMO. The selected performance metric was MSE for the sake of simplicity, defined generally in (1). Regarding Cross Validation (CV), we devoted 80% of samples to training and 20% of samples to testing. Moreover, a Monte Carlo approach was used for the selection of training and test samples. After some stability experiments, results confirmed that 1000 independent runs, T_{iter} , provided robust enough results. The global importance of each variable, GI_T , calculates the mean importance of that variable in T_{iter} , according to (2). Let us define a vector with the GI of all the involved variables, $GI_k = [GI_1, \dots, GI_k]^T$, where $k = \{1, 2, \dots, 122\}$. The performance of the model is assessed similarly, and the analyzed MSE_T is the average of MSE in T_{iter} runs for each value of the selected hyperparameter, and for each number of variables, given by (3) y (4). The difference between $Training\ MSE_T$ and $CV\ MSE_T$ is the upper limit in the second summation, where $N_{training}$ refers to all the samples in EMO, and N_{CV} indicates only the test samples involved in CV.

$$GI_T = \frac{1}{T_{iter}} \sum_{t=1}^{T_{iter}} GI_k \quad (2)$$

$$Training\ MSE_T = \frac{1}{T_{iter}} \frac{1}{N_{training}} \sum_{t=1}^{T_{iter}} \sum_{i=1}^{N_{training}} (y_{o,i} - \widehat{y}_{o,i})^2 \quad (3)$$

$$CV\ MSE_T = \frac{1}{T_{iter}} \frac{1}{N_{CV}} \sum_{t=1}^{T_{iter}} \sum_{i=1}^{N_{CV}} (y_{o,i} - \widehat{y}_{o,i})^2 \quad (4)$$

3. RESULTS

3.1. Training Set Experiments

Training sets may be useful for variable selection frameworks and might avoid using further analysis or more complex techniques. Thus, our first experiment evaluated DTR performance with all the available samples in the training set. The target is the evaluation of the variables' importance. This experiment also uses the Monte Carlo approach, and the outcomes are the mean of T_{iter} independent runs of training. The performance of the DTR model is excellent but too optimistic, as it will be confirmed with CV experiments, because DTR models tend to provide overfitting easily; furthermore, when MaxDepth increases largely as in this case.

Figure 1 shows the performance of the DTR models in this experiment. Within the training set, DTR models become an error-free but only with a high value of MaxDepth, both for arousal (from MaxDepth = 22) and valence (from MaxDepth = 24). Other studies performed good scores for the training sets with linear regression for arousal (MSE = 0.0432) and valence (MSE = 0.1182) [7]. DTR improves those results only for a particular MaxDepth of 3 for arousal and valence.

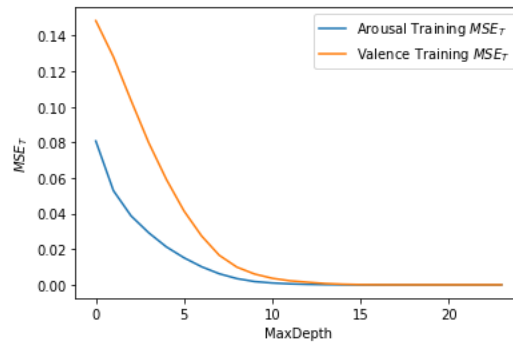


Figure 1 – Training MSE_T of a 122-variables DTR model (Arousal + Valence) vs MaxDepth.

Table 1 presents the main results of this experiment, and it arranges the more relevant variables for the DTR model according to their GI. Perception features based on Loudness became the more significant by far both for arousal and valence. As it has been reported with linear regression experiments in [7], the arousal DTR model seems to require fewer variables than the valence one according to the GI of the 10 most relevant variables. The summation of the GI of 5 first variables in the arousal model results in 0.98, while in the valence model the first 10 variables only reach 0.71.

Table 1 – GI_T of the 10 most relevant variables for the training set of a 122-variables DTR model (Arousal + Valence). Identification of the variables by the name and the position in EMO.

Arousal										
Variable	Loudness_ mean (113)	Fluctuation_ _max (4)	Rms_ mean (1)	Energy_mean (115)	Zerocross_ _mean (6)	Mfcc_mean_ _8 (31)	Rms_std (2)	Mfcc_std_ _9 (45)	Inharmonicity_ std (88)	Hcdf_std (86)
GI_T	0.783	0.052	0.030	0.023	0.0094	0.0045	0.0045	0.040	0.0032	0.0031

Valence										
Variable	Loudness_ std (114)	Chromagram_ _std_2 (102)	Entropy_ std (23)	Chormagram_ _mean_12 (100)	Mfcc_mean_ _5 (28)	Decreaseslope_ mean (3)	Loudness_ _mean (113)	Kurtosis_ std (110)	Energy_ _mean (115)	Brightnes s_mean (10)
GI_T	0.492	0.053	0.044	0.031	0.023	0.020	0.014	0.012	0.011	0.010

3.2. Cross Validation Experiments

A CV procedure provides more robust models than just working the training set. Hence, MSE results within this experiment become worse than in the previous one. As expected, DTR usually overfits in training. Figure 2 shows the MSE decrease in performance for both outputs. Moreover, the models reach a minimum value of MaxDepth and after that, the model overfits; this effect is underlined for valence. The arousal DTR model with CV provides a minimum MSE of 0.0534 and the valence DTR model presents its best MSE of 0.151, with a MaxDepth value of 4 and 3 respectively.

This MaxDepth was close to the training experiment point where DTR improved linear models of previous studies. However, these optimal values are also worse than those of linear regression both for arousal (CV MSE = 0.045) and valence (CV MSE = 0.123) [7], although they are competitive. Some other techniques based on non-linear methods, such as Random Forest, Support Vector Machines or Artificial Neural Networks improved those scores, but they lose interpretability and become complex structures [8], [9], [15].

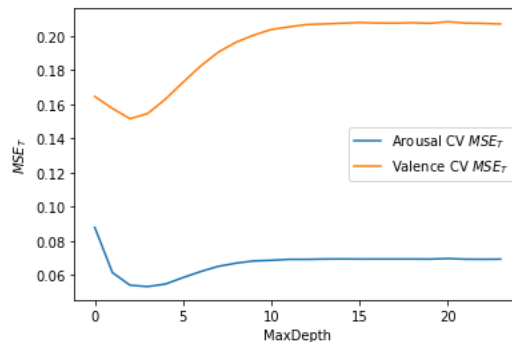


Figure 2 – CV MSE_T of a 122-variables DTR model (Arousal + Valence) vs MaxDepth.

The next step consists of listing the most relevant variables in this experiment. *Table 2* confirms that the pattern of variable importance in the training set appears again when including CV. The most important variable remains the same for both outputs and keeps the large weighting regarding the following relevant variables. There are some differences in the order of the rest of the variables. Thus, the training set experiment might provide enough information about features' importance without an extensive computational load for CV.

In the arousal case, there is only an order change between *Energy_mean* and *Rms_mean*, and in the fifth position the change of *Zerocrossing_mean* by *Mfcc_std_9*. When it comes to valence, some more changes came up in the order and the list of variables. Some features like *Loudness_mean*, *Energy_mean*, *Entropy_std*, *Mfcc_mean_5*, and *Decreaseslope_mean* remained but changed their position. The remainder went out of the list and some new variables were included. Two of these new features were also used in the arousal model.

Table 2 – GI_T of the 10 most relevant variables of a 122-variables DTR model with CV (Arousal + Valence). The variables that kept the training set order are shaded in green, and the ones that changed their position are shaded in blue. In white, new variables regarding the training set.

Arousal

Variable	Loudness_mean (113)	Fluctuation_max (4)	Energy_mean (115)	Rms_mean (1)	Mfcc_std_9 (45)	Energy_std (116)	Zerocross_mean (6)	Flux_mean (50)	Loudness_std (114)	Rms_std (2)
GI_T	0.786	0.053	0.028	0.022	0.008	0.004	0.004	0.003	0.003	0.003

Valence

Variable	Loudness_std (114)	Energy_mean (115)	Rms_mean (1)	Loudness_mean (113)	Flux_mean (50)	Mfcc_mean_5 (28)	Entropy_std (23)	Inharmonicity_mean (87)	Decreaseslope_mean (3)	Fluctuation_max (4)
GI_T	0.478	0.052	0.047	0.036	0.029	0.021	0.014	0.012	0.011	0.009

3.3. Experiments with a Heuristic Filter

The final step of this variable selection proposal is the application of a heuristic filter over the results of the embedded method, which was obtained but the GI of variables when performing the DTR model using all the available samples for training. This dataset shows an extremely relevant variable for both outputs due to their outperformance in GI terms. These variables are the same in the training and CV experiments too.

First, we considered a GI_T summation of the most important variables, X_k , which are greater than 80% in the training outcomes, given by (5). This percentage is a simple *Pareto* rule to check. Hence, we could assess the model with only two variables for arousal and more than ten variables for valence. Furthermore, to collect more information about the variable selection framework, we also performed the DTR models with the two, three, four and five most relevant variables in the training and the CV experiments. For the sake of the readability of the paper, we only scrutinized the results of the arousal output.

$$X_k \in GI_T = \frac{1}{T_{iter}} \sum_{t=1}^{T_{iter}} GI_k \geq 0.8 \quad (5)$$

Thus, according to *Table 1* and *Table 2*, for one and two-variables models for arousal, the training and CV frameworks result in the same features (*Loudness_mean*, and *Loudness_mean + Fluctuation_max*). The rest of the models included some changes in the chosen variables.

Table 3 – DTR models performance after employing the proposed variable selection framework, including a comparison with the initial 122 features model with CV.

Number of Variables	Variable Selection option	Best MSE	MaxDepth of best MSE	MSE difference vs 122 variables model (%)
1	Training / CV	0.07354	3	+ 38.4
2	Training / CV	0.04691	5	-11.7
3	Training	0.04902	5	-7.7
4	Training	0.04891	5	-7.9
5	Training	0.04819	6	-9.2
3	CV	0.04866	5	-8.4
4	CV	0.04868	5	-8.3
5	CV	0.04882	6	-8.1

According to *Table 3*, a reduction of variables led to better models' performance. Models with a few features showed lower MSE than a 122 variables model. Although the more relevant variable features an extraordinary weighting compared to the rest, a one-feature model performs poorly concerning other analyzed alternatives. The heuristic filter seems to work properly, and the enhancement of performance starts at the expected number of variables, which is two. Moreover, the two-features model shows the better score of all the scrutinized models. The use of more variables might generate overfitted models as the increase of MaxDepth did in previous experiments. Thus, these results also suit the general theory of DTR, which shows better generalization as the model remains simple.

The variable selection framework of the training experiments overcame the 122-features model. From three to five features, the training outcomes are slightly worse than the CV experiments regarding MSE, but the complexity of the models is the same with structures between 5 and 6 of MaxDepth.

4. CONCLUSIONS

Regression models require the lowest computational load to save resources from different points of view. Variable selection techniques may help reduce the number of relevant features. The presented variable selection framework provided a dimensionality reduction when DTR model SER from EMO.

DTR revealed a competitive performance as a predictive model for SER. Other approaches overcome MSE, but DTR may improve the interpretability of the model. Moreover, the presented variable selection framework based on DTR provides solutions with few variables, and thus, a parsimonious predictive tool.

The use of all the samples of the dataset for training might help select relevant variables for a DTR model. According to the experiments, the variables' importance in training are similar to those ones obtained by CV. Thus, working with a variable selection framework in the training stage provides a trade-off between computational load and competitive results.

The next research steps focus on the same analysis for valence, the application of this variable selection framework before the training and validation of other regression algorithms, and the performance of this solution counting on more DTR hyperparameters.

ACKNOWLEDGMENTS

The authors acknowledge support by: Agencia Estatal de Investigación - AEI (project SPGRAPH, ref. num. PID2019-105032GB-I00), Comunidad de Madrid and Universidad Rey Juan Carlos (Proyecto I+D Jóvenes Doctores, AUTO-BA-GRAPH, ref. num. F861).

REFERENCES

- [1] P. Lopes, A. Liapis, and G. N. Yannakakis, "Modelling affect for horror soundscapes," *IEEE Trans Affect Comput*, vol. 10, no. 2, pp. 209–222, Apr. 2019, doi: 10.1109/TAFFC.2017.2695460.
- [2] F. Abri, L. F. Gutiérrez, P. Datta, D. R. W. Sears, A. S. Namin, and K. S. Jones, "A Comparative Analysis of Modeling and Predicting Perceived and Induced Emotions in Sonification," *Electronics 2021, Vol. 10, Page 2519*, vol. 10, no. 20, p. 2519, Oct. 2021, doi: 10.3390/ELECTRONICS10202519.
- [3] J. A. Russell, "A circumplex model of affect," *J Pers Soc Psychol*, vol. 39, no. 6, pp. 1161–1178, Dec. 1980, doi: 10.1037/H0077714.
- [4] W. J. Davies, N. S. Bruce, and J. E. Murphy, "Soundscape reproduction and synthesis," *Acta Acustica united with Acustica*, vol. 100, no. 2, pp. 285–292, Mar. 2014, doi: 10.3813/AAA.918708.
- [5] M. Lionello, F. Aletta, and J. Kang, "A systematic review of prediction models for the experience of urban soundscapes," *Applied Acoustics*, 170, Article 107479. (2020), vol. 170, Dec. 2020, doi: 10.1016/J.APACoust.2020.107479.
- [6] J. Fan, M. Thorogood, and P. Pasquier, "Emo-soundscapes: A dataset for soundscape emotion recognition.," *2017 7th International Conference on Affective Computing and*

- Intelligent Interaction, ACII 2017*, vol. 2018-January, pp. 196–201, Jan. 2017, doi: 10.1109/ACII.2017.8273600.
- [7] R. San Millán-Castillo, L. Martino, E. Morgado, and F. Llorente, “An Exhaustive Variable Selection Study for Linear Models of Soundscape Emotions: Rankings and Gibbs Analysis,” *IEEE/ACM Trans Audio Speech Lang Process*, vol. 30, pp. 2460–2474, 2022, doi: 10.1109/TASLP.2022.3192664/MM1.
- [8] S. Ntalampiras, “Emotional quantification of soundscapes by learning between samples,” *Multimed Tools Appl*, vol. 79, no. 41–42, pp. 30387–30395, Nov. 2020, doi: 10.1007/S11042-020-09430-3/FIGURES/4.
- [9] J. Fan, F. Tung, W. Li, and P. Pasquier, “Soundscape Emotion Recognition via Deep Learning,” Jul. 2018, doi: 10.5281/ZENODO.1422589.
- [10] R. San Millán-Castillo, E. Morgado, and R. Goya-Esteban, “On the Use of Decision Tree Regression for Predicting Vibration Frequency Response of Handheld Probes,” *IEEE Sens J*, vol. 20, no. 8, pp. 4120–4130, Apr. 2020, doi: 10.1109/JSEN.2019.2962497.
- [11] R. M. Schafer, *The Soundscape: Our Sonic Environment and the Tuning of the World*. Rochester Destiny Book, 1993. Accessed: Sep. 27, 2022. [Online]. Available: <http://books.google.com/books?hl=en&lr=&id=ltBrAwAAQBAJ&pgis=1>
- [12] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY: Springer New York, 2009. doi: 10.1007/978-0-387-84858-7.
- [13] A. Jović, K. Brkić, and N. Bogunović, “A review of feature selection methods with applications,” *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2015 - Proceedings*, pp. 1200–1205, Jul. 2015, doi: 10.1109/MIPRO.2015.7160458.
- [14] V. Bolón-Canedo, N. Sánchez-Marroño, and A. Alonso-Betanzos, “A review of feature selection methods on synthetic data,” *Knowledge and Information Systems 2012 34:3*, vol. 34, no. 3, pp. 483–519, Mar. 2012, doi: 10.1007/S10115-012-0487-8.
- [15] F. Abri, L. F. Gutiérrez, P. Datta, D. R. W. Sears, A. S. Namin, and K. S. Jones, “A comparative analysis of modeling and predicting perceived and induced emotions in sonification,” *Electronics (Switzerland)*, vol. 10, no. 20, p. 2519, Oct. 2021, doi: 10.3390/ELECTRONICS10202519.

2 - 4 NOVIEMBRE
ELCHE

TECNIACÚSTICA

CENTRE DE CONGRESSOS D'ELX



CONFIRMACIÓN DE INSCRIPCIÓN

Estimado/a Roberto San Millán-Castillo ,

Le confirmamos su inscripción al **Congreso TECNIACÚSTICA 2022**, que se celebrará en Elche, los días 2 al 4 de noviembre de 2022.

Los detalles son los siguientes:

ID Inscripción: #140

Cuotas de Inscripción: No asociado (525 €)

Cuotas de Inscripción: Workshop Clasificación Acústica de Edificios (Acceso libre, previa inscripción) (0 €)

Total pagado: 525€

Acceso a la plataforma:

<https://www.tecniacustica.es/TECNIACUSTICA2022>

Usuario: roberto.sanmillan@urjc.es

Contraseña: Cumpliendo con el Reglamento UE 679/2016, de 27 de abril, General de Protección de Datos, y con la Ley Orgánica 3/2018, de Protección de Datos Personales y Garantía de los Derechos Digitales, no tenemos acceso a su contraseña. Si no la recuerda puede recuperarla pulsando en "¿Ha olvidado su contraseña?", en la caja de inicio de sesión.

Le recordamos que en **Mi Congreso** podrá consultar el estado de su inscripción, comunicaciones enviadas y realizar cualquier otro tipo de actividad relacionada con el congreso.

Atentamente,

SECRETARÍA TÉCNICA

Viajes El Corte Inglés, S.A.

M.I.C.E. Madrid Congressos

Telf: (+34) 91 330 07 55

tecniacustica@viajeseci.es

VIAJES El Corte Inglés

CONGRESOS CIENTÍFICO-MÉDICOS