Music Recommendation System for Human Attention Modulation by Facial Recognition on a driving task: A Proof of Concept

Roberto Avila - Vázquez[1], Sergio Navarro – Tuch[1], Rogelio Bustamante – Bello[1], Ricardo A. Ramírez Mendoza[1] and Javier Izquierdo - Reyes[1]

[1]*Tecnológico de Monterrey, School of Science and Engineering, Calle del Puente 222, Ejidos de Huipulco, 14380, Tlalpan, Ciudad de México, Mexico*

Abstract. The role of music on driving process had been discussed in the context of driver assistance as an element of security and comfort. Throughout this document, we present the development of an audio recommender system for the use by drivers, based on facial expression analysis. This recommendation system has the objective of increasing the attention of the driver by the election of specific music pieces. For this pilot study, we start presenting an introduction to audio recommender systems and a brief explanation of the function of our facial expression analysis system. During the driving course the subjects (seven participants between 19 and 25 years old) are stimulated with a chosen group of audio compositions and their facial expressions are captured via a camera mounted in the car's dashboard. Once the videos were captured and recollected, we proceeded to analyse them using the FACET™ module of the biometric capture platform iMotions™. This software provides us with the expression analysis of the subjects. Analysed data is postprocessed and the data obtained were modelled on a quadratic surface that was optimized based on the known cestrum and tempo of the songs and the average evidence of emotion. The results showed very different optimal points for each subject, that indicates different type of music for optimizing driving attention. This work is a first step for obtaining a music recommendation system capable to modulate subject attention while driving.

1 Introduction

The role of music listening in the vehicle while driving represents an important element of the design of the user experience. Ünal [1] suggests that the presence of music selected by the participant while driving increases the accuracy of lateral control and better time response latency on speed change. Other researchers like Brodsky [2] argues that music has a positive influence in the emotions of the driver, but the effect on driver accuracy is still unclear. Mizoguchi [3] stablishes that music for the driver with slow tempo and medium volume level could work as an effective safety driver assistance. Also, the use of music could reduce the cardiovascular damage of driving by the induction of positive emotions [4]. Unfortunately, the election of soundtracks lacks of a formal methodology, thus is necessary an improvement in order to obtain better and reproducible results in this kind of experiments.

Some procedures have been proposed on the election of music as a safety drive assistance or a comfort element. Kinoshita [5] describes a system based on the evaluation of a visual simulated scenario for the election of a playlist with a specific musical genre. This idea was improved by Krishnan [6], that include the metadata of the song, the musical features and demographic data of the driver. This process of recommendation could be done by different approaches, like a distance of subject punctuation in the article cited above, a self – organizing map [7] or an artificial neural network [8]. The usability and specificity of the variables measured depends on the state of the art in the instrumentation and research on physiological and psychological values.

This work proposes a system that relates the soundtrack listened by the driver with the attention of the driver in the road. Previous work [9] suggests that the number of emotions expressed by a subject is related with the attention level of the subject in the task. Thus, a decrease in the number of emotions detected while listening to music could be used an approximation of the effectiveness of the soundtrack heard by the user as an attention modulator. For this purpose, iMotions FACET module [10] is used for emotion recognition by facial expression.

2 Methodology

For this study, a group of young drivers was recruited. The selected age range was chosen cause it's one of the groups with most accidents related with euphoric driving and distraction in which the emotions of the subject have an important role. Seven volunteers between 19 and 25 years old (2 women and 5 men) all of them with valid

driver's license and with knowledge in driving automatic transmission vehicles. All participants were invited to be part of this study and none of them was against him/her will. The subjects were labelled from S1 to S7 for analysis maters. The circuit selected for the driving task is presented in Figure 1. The circuit had a length of approximate 3.5 kilometres with a normal traffic of a scholar zone.

Figure 1. Driving circuit used in the experiment

The experiment was designed with a fixed duration of six minutes and 45 seconds. Six songs of 45 seconds were selected from a free music database [11] characterized by the emotional reaction. This duration was elected for an average speed of 35 km/h and an adequate duration of the test without an effect of fatigue that could affect the results.

Mean tempo (Te) [12] and cepstrum (Ce) were selected as significant characteristics for music classification and relevant impact while driving. The characteristics of these two features in the whole database are presented in Table 1 and the properties ones of the songs selected appeared in Table 2. These songs allow to have an adequate sample of the database with an approximate distribution of tempo and cepstrum that could be used in a response surface model described in Equation 1 [13].

Table 1. Ranges of Cepstrum and Tempo of free music database

Variable	Minimum	Maximum
Cepstrum	3876.44	28718.95
Tempo	52.14	159.94

The ranges of cepstrum and tempo indicates to us a database with high variability of genre in the content of the database. This variability could be used to have an adequate surface response representative of subject behaviour.

The playlist was designed in the same order as it is presented in the table, with a silence of 1 minute as a baseline and 15 seconds of silence between each song.

Table 2. Ranges of Cepstrum and Tempo of songs selected

Song	Cepstrum	Tempo
167.mp3	11126.61	103.19
389.mp3	4453.76	80.98
562.mp3	6972.56	159.67
649.mp3	7301.37	60.79
797.mp3	9307.82	123.96
852.mp3	11864.08	159.94

The driver's face was recorded with a webcam connected to a computer controlled by one researcher that supervises the proof. Also, the computer reproduces the playlist with its loudspeakers. The subject receives oral instructions about the objective of the experiment and the course of the driving circuit. The record was stopped after six minutes and 45 seconds.

iMotions system through the implementation of the FACET module. Such module analyses the face images in order to detect the movement by tension or relaxation of the muscles, identifying the Action Units (AU) after the detection of such units the correlation of the AU by the use of the Facial Action Coding System (FACS) deploys a value related with the probability of the emotion being displayed by the subject. In this test we worked with the evidence records of 9 emotions (joy, anger, disgust, contempt, sadness, surprise, fear, confusion and frustration). These emotions were measured with facial recognition with a rate of 15 samples per second. Emotions presented were counted and the signal obtained were smoothed with a mean filter of 200 samples. The maximum of every period of song were registered and a quadratic model was modelled as is presented in Equation 1.

$$S = p_{00} + p_{10} * Ce + p_{01} * Te + p_{20} * Ce^2 + p_{11} * Ce * Te + p_{02} * Te^2 \quad \text{Eq. 1}$$

Symbol	Variable
S	Average Emotion Predicted
p_{nn}	Weigths
Ce	Cepstrum
Te	Tempo

This model is a common assumption in experimental design without a known response, thus is a preliminary model that could be used to predict attention related to audio features. This model is relevant because includes the interaction between both features, a hypothesis that not had been reported in literature. After the

determination of this surface each surface, was optimized in order to obtain the optimal sound characteristics that produces the maximum attention for each subject. This maximization process was obtained with a constrain of tempo between 60 and 160 beats per minute and a mean cepstrum between 5000 and 12000.

3 Results

Figure 2 shows the signal obtained from its number of emotion in the whole proof. The segment while the songs were reproduced are in darker grey and silence in lighter grey.

As we see in Figure 2, the behaviour of the signal along time is not homogeneous. These differences indicate also a variability in the number of emotions expressed across the driving task. This difference could be associated to environmental conditions, but the differences founded in every subject indicates a personal preference to music because every subject have similar driving conditions.

The higher value of the signal post processed (that indicates a major number of emotions expressed in that point) for every subject is summarized in Table 3.

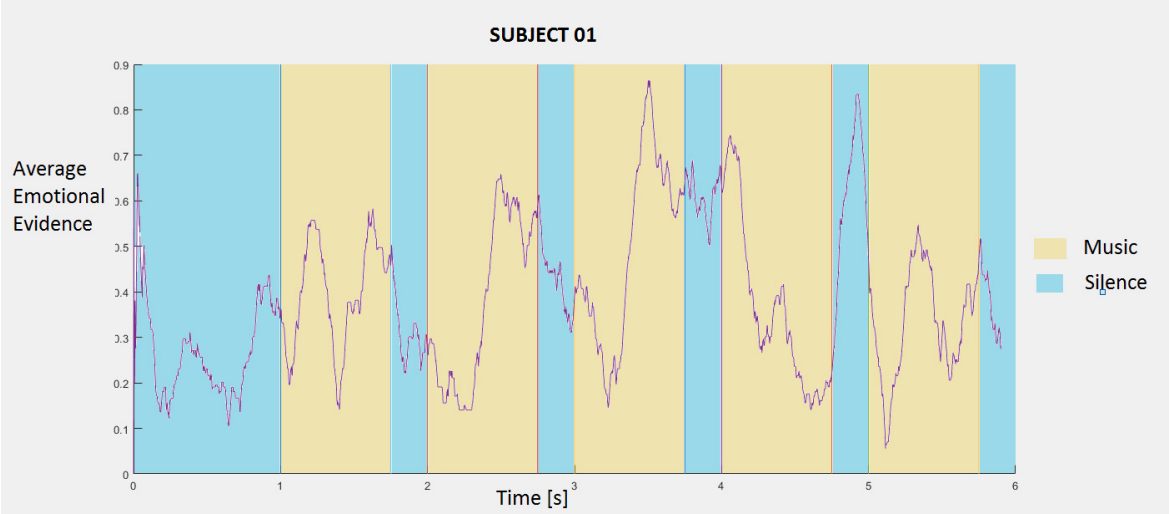

Figure 2. Example of a subject signal of time vs. number of emotions.

Table 3. Maximum number of emotions per subject per song.

Song	S1	S2	S3	S4	S5	S6	S7
167.mp3	0.58	0.75	1.27	1.347	1.77	1.29	0.26
389.mp3	0.65	1.76	1.35	1.28	2.135	1.35	0.35
562.mp3	0.865	1.91	1.51	1.21	2.28	0.995	0.48
649.mp3	0.75	2.3	0.95	1.855	1.525	1.2	0.49
797.mp3	0.83	2.27	1.45	1.38	1.595	1.15	0.62
852.mp3	0.515	1.32	0.31	1.35	1.3	1.08	0.53

The surfaces described in the methodology were obtained for each subject. The model elected in the methodology allows to find a maximum value that indicates an optimal induction of attention. An example of this kind of surfaces is presented in Figure 3. Unfortunately, the parameters for each surface, presented in Table 4, indicates strong differences in each model and optimal values. These results could indicate important differences from each subject, but also problems in the acquisition of each driving test. The parameters obtained for each surface are summarized in Table 4. Some values of the model are small. That situation could indicate a simplification of the equation proposed, but also affects the optimal values for each subject. Thus, its necessary a research on the meaning of each parameter of the model in order to obtain adequate musical features for the creation of playlists.

Information in Table 5 reflexes the high diversity of optimal characteristics as we expected in the diversity of parameters founded in Table 4. The optimal parameters are inside the values of database ranks; thus its values are constrained in typical values of songs. This optimization could be used in the election of individualized playlists, but its efficacy must be proven in future experiments.

Table 4. Parameters of the quadratic forms of tempo and cepstrum vs Average number of emotions

Parameter	S1	S2	S3	S4	S5	S6	S7
p00	-6.04E-01	-6.84E-01	-3.10E+00	1.991	1.568	1.542	-2.57E-01
p10	2.92E-04	1.20E-03	3.01E-04	3.03E-04	-3.30E-04	-1.61E-04	2.75E-04
p01	5.95E-03	-2.47E-02	6.79E-02	-3.06E-02	3.15E-02	7.75E-03	-5.85E-03
p20	-2.14E-08	-1.29E-07	-7.73E-09	-2.25E-08	3.96E-08	1.10E-08	-3.09E-08
p11	2.49E-07	6.90E-06	-2.50E-06	9.31E-07	-3.86E-06	-1.74E-07	2.00E-06
p02	-2.96E-05	-1.27E-04	-2.02E-04	8.02E-05	1.37E-05	-3.93E-05	-3.79E-05

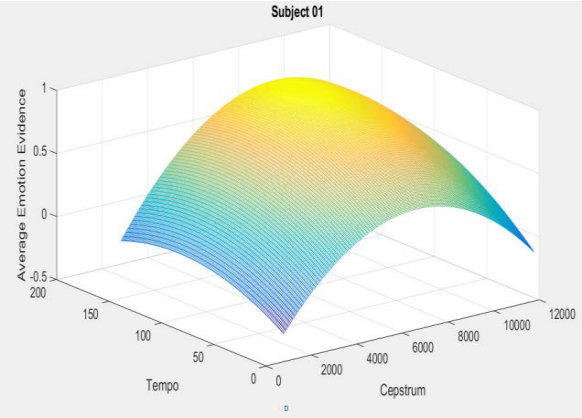

Figure 3. Subject response modelled as a quadratic model

Table 5. Optimal characteristics of songs that maximizes attention

Subject	Optimal Cepstrum	Optimal Tempo
S1	6.47E+03	6.00E+01
S2	7.46E+03	1.05E+02
S3	5.00E+03	1.37E+02
S4	7.97E+03	6.00E+01
S5	1.20E+04	1.60E+02
S6	5.00E+03	8.76E+01
S7	9.62E+03	1.60E+02

4 Discussion and Analysis

This experiment shows an important area of exploration in driving ergonomics. Music playlist could be understood as an active element in the driver's environment, that allows to improve the driver's performance. This election must be transformed from a subjective election of songs to an objective evaluation of the subject based on the acoustic characteristics of the samples. As we see in the present work, it is possible to reduce the experiment to some songs elected in a way that allows to make a model of the subject response and finally a prediction about the kind of songs that could be adequate for every subject.

Table 5 is an interesting example about how individual preferences affects the behaviour of the driver. We were not able to identify a pattern that indicates that users with similar demographics also shares similar music preferences. This methodology allows us to identify this differences and approximates a first results about how to identify optimal characteristics of the songs that must be reproduced in a driving task. A preliminary analysis of the results obtained confirms previous results [2] about the preference of young people of high tempo and aggressive music, that could be associated with lower cepstrum and electronic music compared with acoustic music with higher cepstrum. Not all the electronic music had low cepstrum and other genres could have these characteristics, so the model must be refined in order to obtain an adequate music playlist for driver's attention optimization.

This experiment also was performed on realistic conditions. That implies a problem in face recognition because the variation of sun light, road conditions and obstacles in the way. Ideal conditions could imply a best convergence of the results in recognizable patterns of optimal sound condition, but also could affect the analysis of real driving tasks. This equilibrium and the differences in the subject in both experimental environments must be deeply analysed to adequately extrapolate conclusions between real and simulated driving.

This work presents a first approximation about the relationship between attention, emotion expression and music influence on driving task. The use of a quadratic surface in the optimization method is a valid first approximation, but a real approximation about the significance of the surface must be explored in future work. The mathematical relationship between audio characteristics and the human response could indicate behavioural mechanisms that could not be described yet. Also the inclusion of other audio parameters could make a more complicate and complete model about these responses and give ideas about hidden mechanism of daily life activities.

5 Conclusions

This work shows a method for music playlist creation based on the individual characteristics of the driver and the driving task. Cesptrum and tempo proved to be valuable indicators for music classification and playlist optimization, but other parameters must be included on future experiments. Quadratic model of music features vs. subject response works as a valuable starting point for the evaluation of the subject state, particularly attention while driving. These discoveries must be confirmed by other experiments that include other physiological values. In future work, other sensors can be included to analyse the correlation between the variables mentioned in this work and the impact on the dynamics of the car. An example of future works could include accelerometers measurements in the circuit to correlate the behaviour of the driver and the emotions affected by the auditory stimuli.

References

1. A.B. Ünal, D. de Waard, K. Epstude & L. Steg, *Transportation Research Part F: Traffic Psychology and Behavior* **21**, 52-65 (2013).
2. W. Brodsky & M. Kizner, *Transportation Research Part F: Traffic Psychology and Behavior* **15 (2)**, 162-173, (2012).
3. K. Mizoguchi & S. Tsugawa *Vehicular Electronics and Safety (ICVES), 2012 IEEE International Conference on,* IEEE pp. 117-121 (2012,July)
4. S.H. Faiclough, M. van der Zwaag, E. Spiridon & J. Westerink *Physiology & Behavior* **129**, 173-180 (2014)
5. Y. Kinoshita, Y. Masaki, T. Muto, K. Ozawa & T. Ise *Consumer Electronics, 2009. ICE'09. IEEE 13th International Symposium on* , IEEE, pp. 94-98 (2009, May).
6. A.S. Krishnan, X. Hu, J.Q. Deng, R. Wang, M. Liang, C. Zhu, Y.K. Kwok *IEEE 7th International Conference on Cloud Computing Technology and Science (CloudCom)* (2015)
7. N.H. Liu *EURASIP Journal on Audio, Speech and Music Processing* **20 (1)** 52-65 (2013)
8. N.H. Liu *Multimedia tools and applications* **72 (2)**, 1341-1361
9. R. Avila – Vázquez, R. Bustamante – Bello, R.A. Ramírez – Mendoza, A. Beltrán Fernández, J. Izquierdo Reyes, S.A. Navarro – Tuch *10th International Conference on Advanced Computational Engineering and Experimenting ACE-X* (2016, July).
10. iMotions. iMotions Emotient https://imotions.com/blog/facial-action-coding-system/
11. M. Soleymani, M.N. Caro, E.M. Schmidt, C.Y. Sha, Y.H. Yang *Proceedings of the 2nd ACM International Workshop on Crowdsourcing for Multimedia* pp. 1-6 (2013, October)
12. W. Brodsky *Transportation Research part F: Traffic Psychology and Behavior* **4 (4)**, 219 – 241 (2001)
13. K. Hinkelman, O. Kempthorne *Design and Analysis of Experiments* p.504 (2007)