Audio Engineering Society

Convention Paper 8812

Presented at the 134th Convention
2013 May 4–7 Rome, Italy

This Convention paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This convention paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Investigating Auditory Human-Machine Interaction: Analysis and Classification of Sounds Commonly Used by Consumer Devices

Konstantinos Drossos[1], Rigas Kotsakis[2], Panos Pappas[3], George Kalliris[2] and Andreas Floros[1]

[1]*Digital Audio Processing & Applications Group, Dept. of Audiovisual Arts, Ionian University, Corfu, Greece*

[2]*Laboratory of Electronic Media, Dept. of Journalism and Mass Communication, Aristotle University of Thessaloniki, Thessaloniki, Greece*

[3]*Dept. of Sound & Musical Instruments Technology, Technological Educational Institute of Ionian Islands, Lixouri, Greece*

Correspondence should be addressed to Konstantinos Drossos (kdrosos@ionio.gr)

ABSTRACT

Many common consumer devices use a short sound indication for declaring various modes of their functionality, such as the start and the end of their operation. This is likely to result in an instantaneous and intuitive auditory human-machine interaction, imputing a semantic content to the sounds used. In this work we investigate sound patterns mapped to "Start" and "End" of operation manifestations and explore the possibility such semantics' perception to be based either on users prior auditory training or on sound patterns that naturally convey appropriate information. To this aim, listening and machine learning tests were conducted. The obtained results indicate a strong relation between acoustic cues and semantics along with no need of prior knowledge for message conveyance.

1. INTRODUCTION

Nowadays, most consumer electronic devices (such as personal computers, mobile phones, hand held cameras, etc) use sound as an extra path of Human-Machine Interaction (HMI). The human user of such appliances usually expects an audio event as a response to his actions. This notion is more prominent and very typical during the start or the end of a device's operation session, where the sound manifestation of the functionality's mode changing (FMC), i.e. switching on or off the device, can instantaneously inform the user.

Different manufacturers have employed various sound patterns for the aforementioned intercourse. Nevertheless, it is widely recognized that users can correctly, instantaneously and intuitively understand the auditory message despite the variance of motifs. In this work we investigate whether the correct perception of the auditory message is based on the listener's prior training, occurred by his antecedent interaction with a specific device, or it is naturally occurred, i.e. the user can understand the auditory message without any prior knowledge. To this aim, we combine listening and machine learning tests on short sounds recorded from various consumer devices. This coalition intends to evaluate the accuracy of the auditory message conveyance to humans and to examine any common technical characteristics in sound patterns used for the same message delivery.

To this aim, we hereby examine sounds employed for the "Switch ON" and "Switch OFF" FMC manifestations. These categories appear to be widely and frequently used in HMI along with sounds indicating danger and faults in the operation of a device. Moreover, it is likely that many users of various appliances tend to recognize the latter from its FMC manifestations, with most prominent the ones taken into account in this work.

The rest of the paper is organized as follows: Section 2 includes a brief presentation of existing researches in audio-related HMI and semantic analysis. Section 3 describes the experimental procedure followed within the scope of this work. The results obtained are analytically presented and discussed in Section 4. Finally, Section 5 concludes the work and defines a number of issues that may extend the outcome of this work and can be considered in the future.

2. RELATED RESEARCH

To the best of authors' knowledge, there are no bibliographic references regarding the intuitive conveyance of messages through the sound channel in HMI. However, there are numerous researches in various alternative and / or legacy fields, like Auditory Displays (AD), Music Information Retrieval (MIR), Audio Emotion Recognition (AER) and Audio Semantic Analysis, all concerned with the perception of sound, in different of course perspectives.

More specifically, in AD there are various investigations concerning the delineation of stimuli with sound [1, 2]. The latter is performed by mapping the under-consideration information to sound's technical characteristics [1] or by adopting patterns, like frequency continuous decrease [2], in order to represent interactions and information (i.e. a decrement in size). Also, the correlation of acoustic cues with high level information is a fact very frequently considered by MIR. In this discipline, technical features are used in conjunction with machine learning techniques in order to classify the former according to notional attributes, like genre [3, 4]. The classification/categorization of audio in accordance to conceptual aspects is also examined by AER, focusing mainly on the recognition of emotion in sound, usually using the same procedure as in MIR field [5, 6, 7].

Typical features employed in the above disciplines include, but are not limited to, the ones listed in Table 1.

Table 1: Technical features commonly used in audio classification tasks

RMS energy	Low energy	Tempo
Beat spectrum	Onsets	Mode
Roll off	Zero crossing	Brightness

Based on the above technical characteristics of the sound signal, various researches in the aforementioned fields have achieved confident accuracy results for audio classification and/or categorization [3, 5, 8] emotion recognition from music [5] and semantic analysis [8]. Moreover, considering some of the above extracted features, a typical audio classification process generally consists of two stages: a) training of the classification system, and b) classification using an appropriate algorithm. Typical classification algorithms include Support Vector Machine, Linear Regression, Artificial Neural Networks implementations (like Artificial Neural Systems), Gaussian Mixture Models and others [3, 9, 8].

In the former stage, the selected algorithm is trained using a ground truth data set. This data set contains not only the extracted features, but also the annotation of the classes used for the categorization. The efficiency of both features and algorithm chosen is then tested using a test data set for the classification task. The latter data set consists by particulars not included in the ground truth data set. According to recent researches [9, 10], the same methods and procedures can be used for non-musical audio signals, such as environmental sounds or, evidently, for sounds targeted to human-machine interaction.

Due to the aforesaid scarcity of existing works on the evaluation of the possibilities for an intuitive conveyance of messages in HMI and on the other hand, the requirement for investigating wether sounds used to communicate congruent messages exhibit similarities in their acoustic cues, in this work methods and techniques from the above fields were employed. Hence, extraction of technical features, including those commonly used in the aforementioned fields was performed, combined with well-established and frequently-used classification/categorization algorithms, engaged also in all the above disciplines.

Typical sound waveforms mapped to “Switch ON” and “Switch OFF” FMC manifestations were employed, since they are subject to differences between different equipment manufacturers, designed according to their preferences on musical motifs. In addition, these specific audio FMC manifestation are not related to sounds commonly used for “symbolic” representations, as those mentioned in [2].

3. EXPERIMENTAL SEQUENCE

In order to examine common technical characteristics of sounds mapped to same FMC manifestations and consequently evaluate the accuracy of audio HMI messages’ conveyance, we combined the results obtained from well-established methods from the fields mentioned in Section 2 with listening tests respectively. Towards this aim, we recorded a set of audio stimuli, pre-processed them, extracted features from the audio data set and finally performed a series of tests. The latter consist of actual listening tests, using the audio data set employed, and machine learning tests, using the set of the extracted features.

The data set used consisted of 27 sounds from “Switch ON” and 14 from “Switch OFF” FMC manifestation, leading to a total of 41 audio waveforms. For both categories, sounds from appliances as well as software applications were considered. Table 2 illustrates a representative subset of the recorded sounds with respect to their FMC manifestation.

Table 2: Part of the recorded sounds used

Recorded Sound	**FMC Manifestation**
Video camera	Switch On & Off
Camera	Switch On
Mobile phone	Switch On & Off
PC Operating System	Switch On & Off
Video game console	Switch On
Car	Switch Off
Television	Switch On & Off

All sound recordings were performed using a handheld digital recorder with $44.1\,kHz$ sampling frequency and $16\,bits$ resolution. Their average time length was $3,9$ seconds, ranging from $0,2$ to a maximum of $18,9$ seconds. All sounds can be found at: https://www.dropbox.com/s/hlp8rq3m7tqlo1k/sounds.zip.

All the recorded sounds were initially pre-processed. At this stage, the sound samples were firstly normalized, in order to eliminate potential differences in the corresponding sound pressure level, and decomposed in frames of 0.01 seconds length with 20% overlap prior to exploiting rapidly changing attributes. Subsequently, for each of the recorded sounds, the set of features presented in Table 3 was extracted for each generated frame. For the feature extraction process, the widely-employed MIR Toolbox in MATLAB was used [11].

This process led to a set of values per frame for each feature and for each sound. From this collection, a compendium of statistic measures was calculated for each feature, where applicable, in order to describe its values’ variance for successive frames. The result was a total set of 230 features. The statistic measures used are the following:

- Mean
- Standard Deviation (Std)
- Flatness

Table 3: The extracted features for each frame, based on the categorisation of MIR Toolbox [11]

Rhythm Related	
Onsets	Fluctuation
Beat Spectrum	Event Density
Pulse Clarity	
Energy Related	
RMS energy	Low Energy
Timbre Related	
Attack Time	Attack Slope
Zero Crossing	Roll Off
Brightness	Roughness
Regularity	
Pitch Related	
Inharmonicity	
Tonality Related	
Chromagram	Key
Mode	Harmonic Change Detection Function
Structure Related	
Novelty	
Waveform Related	
Centroid	Spread
Flatness	Kurtosis
Entropy	Skewness

- Spread
- Kurtosis
- Skewness
- Slope
- Centroid

3.1. Listening Tests

As mentioned previously, the evaluation of intuitive conveyance of messages in HMI was performed with listening tests. These were conducted in a typical office enclosure and during common working hours, in order to simulate an adequate range of typical devices' usage environmental conditions. Sound reproduction was performed through a common portable computer. This allowed to realize audio reproduction over a short-distant loudspeaker setup, which is typical for the majority of common consumer devices.

During the tests, two different groups of participants were considered. The first one was formed by human subjects that could recall using an appliance of any kind (or software) with “Switch ON” or “Switch OFF” FMC manifestation. The second one included those that could not. Regarding the latter group, participants over 70 years old from the county of Preveza - Greece, were chosen in order to eliminate the possibility for any prior experience with appliances that demonstrate their FMC using the audio channel. In addition, in the case that a participant from the second group could actually recall any of the sounds used in the listening tests, he was automatically excluded from the test process. The resulting total“Switch ON” set of participants was 54 with 70,4 % (i.e. 38) of them belonging to the first group and 29,6 % (i.e. 16) to the second one.

Each participant had to listen to each sound and provide his personal annotations of the perceived FMC. For this process, the complete sound set was used and if a human subject could recall the sound, then the annotation for that particular stimulus was skipped (while it was not valid at all for all the participants belonging into the second group, as noted previously). This led to a total of 1989 annotations, for all stimuli considered. From those, 67 % (i.e. a total number of 1333) were obtained from participants belonging to the first group, whereas the rest 33 % (i.e. 656 annotations) from the second group. In addition, from the total set of annotations, 65,5 % (1301) corresponded to the “Switch ON” and 35,5 % (688) to the “Switch OFF” FMC manifestation.

Annotation was performed in a “Question and Answer” basis after the reproduction of each audio stimulus. Each human subject was asked if the reproduced sound was conveying a “Switch ON” or “Switch OFF” FMC message. Each answer was recorded along with the sound that was reproduced and the group in which that particular subject belonged.

The group distribution of “Switch ON” FMC manifestation annotations was 66,8 % (i.e. 869 annotations) and 33,2 % (432) for the first and second group respectively. Accordingly, the “Switch OFF” FMC manifestation resulted into 67,4 % (i.e. 464)

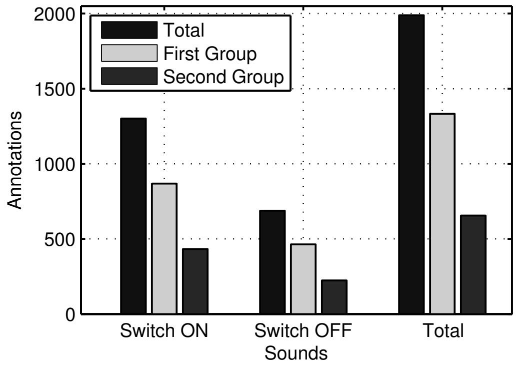

Fig. 1: The annotations of sounds per group and in total

and 32, 6 % (i.e. 224) for the first and second group respectively. The above distribution of annotations is summarized in Figure 1.

The obtained participants' annotations were compared to the intended FMC manifestation of the audio stimulus and the evaluation of annotations' accuracy performed for each group separately and in total, for both FMC manifestations. The comparison was expressed in terms of the ratio, in percentage, of the correctly annotated sounds and the total sounds for each FMC manifestation. Moreover, the average accuracy for both cases ("Switch ON" and "Switch OFF" FMC) was calculated.

3.2. Machine Learning Tests

During the machine learning tests, initially a feature evaluation process was conducted followed by the corresponding classification task. The evaluation of the 230 features set implemented using two different ranking algorithms in the Weka environment, namely the "InfoGainAttributeEval" and the "OneRAttributeEval" ones [12]. The former evaluates the importance of each attribute separately by estimating the information gain, with respect to the class, using entropy metrics. The second one is a simple classification technique that generates, for each feature individually, an one-level decision tree and accounts the number of corrected classified instances [13, 14].

Table 4: First 50 salient features as evaluated with "InfoGainAttributeEval" Weka algorithm. "Chrom." stands for of "Chromagram". First features in the Table are ranked higher and ranking is per row.

InfoGain	
HCDF Flatness	Chrom. 11 Skewness
Chrom. 12 Skewness	Mode Skewness
Chrom. 11 Spread	Chrom. 12 Kurtosis
Novelty Spread	Chrom. 11 Kurtosis
Chrom. 2 Kurtosis	Inharmonicity Mean
Regularity Spread	Chrom. 1 Skewness
Chrom. 1 Centroid	Chrom. 1 Flatness
Inharmonicity Std	Chrom. 1 Kurtosis
Inharmonicity Slope	Roughness Skewness
Roughness Centroid	Roughness Flatness
Roughness Std	Roughness Mean
Roughness Kurtosis	Roughness Slope
Regularity Skewness	Regularity Kurtosis
Regularity Centroid	Regularity Flatness
Regularity Mean	Roughness Spread
Regularity Slope	Regularity Std
Chrom. 5 Flatness	Chrom. 5 Spread
Chrom. 5 Centroid	Chrom. 4 Spread
Chrom. 5 Skewness	Chrom. 5 Kurtosis
Chrom. 6 Spread	Chrom. 7 Skewness
Chrom. 7 Kurtosis	Chrom. 6 Skewness
Chrom. 6 Kurtosis	Chrom. 6 Centroid
Chrom. 6 Flatness	Chrom. 2 Spread
Chrom. 3 Skewness	Chromagram 3 Kurtosis
Chrom. 2 Skewness	Chromagram 1 Spread

In Tables 4 and 5 is the ranking of features that prevailed by testing the whole training set with the above evaluation algorithms. Several features appear in both Tables but in different ranking. Many feature selection processes were conducted, concluding in the fact that the OneRAttribute algorithm offers a more reliable estimation in the evaluation process, because it tests the respective classification performance of each individual feature instead of the partial Information Gain obtained from the InfoGain algorithm. Thus the ranking results of OneRAttribute evaluation are exploited as an indicative basis for the formulation of the final salient feature vector without excluding the ranking of the InfoGainAtrribute algorithm.

Although some classification techniques, like artifi-

Table 5: First 50 salient features as evaluated with "OneRAttributeEval" Weka algorithm. "Chrom." stands for of "Chromagram". First features in the Table are ranked higher and ranking is per row.

OneR	
Attack Time Slope	Chrom. 1 Slope
Brightness Spread	Roll Off Std
Regularity Slope	Chrom. 4 Mean
HCDF Flatness	Beat Spectrum Skewness
Chrom. 2 Flatness	Beat Spectrum Slope
Zero Cross Mean	Mode Slope
Entropy Std	Brightness Centroid
Novelty Std	Beat Spectrum Mean
Centroid Std	Chrom. 7 Centroid
Chrom. 8 Slope	Zero Cross Slope
Chrom. 3 Mean	Chrom. 2 Centroid
Chrom. 10 Centroid	Chrom. 1 Std
Zero Cross Spread	Regularity Centroid
Chrom. 4 Flatness	Chrom. 9 Mean
Chrom. 1 Centroid	Beat Spectrum Kurtosis
Beat Spectrum Std	Novelty Flatness
Spread Mean	Chrom. 11 Mean
Chrom. 9 Std	Chrom. 9 Slope
Flatness Std	Chrom. 11 Slope
Brightness Mean	Attack Slope Centroid
Chrom. 12 Slope	Key Slope
Chrom. 5 Mean	HCDF Std
Zero Cross Flatness	Entropy Slope
Inharmonicity Std	Roughness Spread
Regularity Skewness	Spread Std

cial neural network training, may reject the correlated features by adjusting the inner weights of the representative nodes, other algorithms, like regressions, are quite sensitive in the feature correlation subject. Therefore, in order to evaluate the feature dependencies and correlations, a Principal Component Analysis (PCA) was conducted, before proceeding in training and classification experiments. This revealed that correlations appeared in low-ranking features which, consequently, had to be excluded while forming the final salient feature set for the classification step. In addition, the most useful result, obtained from PCA, refers to the revealed ranking order of 27. Therefore, the final feature vector that has been formulated, taking also into consideration the ranking order of the evaluation algorithms, is presented in Table 6.

For the classification task, three different training algorithms were employed, namely Artificial Neural Systems (ANS), Logistic Regressions and Sequential Minimal Optimization (SMO).

ANS network topology included two sigmoid hidden layers and a linear output layer, providing the capability of achieving the formulation of efficient generalization rules and conclusions [13, 14, 15, 16]. Logistic Regression was used due to its ability of structuring a non-linear regression model that relates the classification decision to the output probability result [17] and SMO as an iterative Support Vector Machine technique. The construction of the classification models with the used algorithms implemented using the k-fold validation technique, i.e. for k iterations, the whole set of instances divided in k subsets, and using k-1 subsets for training purposes and the remaining set for testing the developed model [13, 14].

Since the total number of the audio dataset is 41 (prime number), the selected number of folds were $k = 7$ and $k = 3$ which are the nearest integer multiples of 42. The 7-fold validation offers the balanced division of the input instances ($6 \times 7 = 42$) while 3-fold validation favors the generalization potentials of the classification scheme with limited number of iterations. Finally, 41-fold validation (Leave-One-Out/One-Out technique) was employed, in order to utilize the maximum number of input samples in the process of developing the classification model.

Accuracy evaluation for the used algorithms was performed using Performance/Recognition Rating (PR). This is defined as the ratio, in percentage, of the number of correctly classified instances (Ncc) to the total number of input instances (N) and is calculated as shown in Equation 1 [13].

$$PR = \frac{N_{cc}}{N} \times 100 \tag{1}$$

Also, the partial recognition rate in each class (PR_{ci}) was measured. It is defined as the ratio, in percentage, of the number of correctly classified instances in class i (N_{cci}) to the total number of instances assigned in the respective class (N_c) and is

Table 6: The final features used in machine learning tests

No.	Feature	No.	Feature
1	HCDF Flatness	2	Novelty Std
3	Entropy Std	4	Roll Off Std
5	Regularity Slope	6	Chromagram 4 Mean
7	Attack Time Slope	8	Beat Spectrum Skewness
9	Chromagram 2 Flatness	10	Beat Spectrum Slope
11	Zero Cross Mean	12	Mode Slope
13	Brightness Spread	14	Brightness Centroid
15	Chromagram 1 Slope	16	Beat Spectrum Mean
17	Centroid Std	18	Chromagram 7 Centroid
19	Chromagram 8 Slope	20	Zero Cross Slope
21	Chromagram 3 Mean	22	Chromagram 2 Centroid
23	Chromagram 10 Centroid	24	Chromagram 1 Std
25	Zero Cross Spread	26	Regularity Centroid
27	Chromagram 4 Flatness	28	Chromagram 9 Mean

defined in Equation 2 [13]. PR and PR_{ci} are derived from the confusion matrices at the end of classification process [13].

$$PR = \frac{N_{cci}}{N_{ci}} \times 100 \qquad (2)$$

4. RESULTS & DISCUSSION

A summary of the listening and the machine learning test results is shown in Table 7 and 8 respectively. A prominent result is that the subjective listening evaluation performed worse than machine learning. Participants' scores are lower than the classification algorithms outcomes for both FMC manifestations considered in this work. In addition, human annotation presented an inability in correctly classifying sounds to the "Switch OFF" FMC. Regarding the "Switch ON" category, the highest score was observed in the second group of subjects with a value of 61, 57 %. Also, both groups seem to correctly classify the data set with an overall accuracy percentage of 61, 42 %. In the case of the "Switch OFF" FMC manifestation the corresponding overall rating was below 50 % and the highest score was achieved by the first group with a value of 51, 07 %.

The overall classification performance from ANS, Logistic Regression and SMO algorithms was high and varied from the lowest value of 85, 37 % with Logistic Regression to the highest value of 97, 56 % when ANS was employed. It has to be noted that ANS retain the comparative advantage in overall discrimination rates, independent of the number of respective input sets folds for each training algorithm. SMO approached the classification performances of ANS, while Logistic Regression decreased the overall discrimination rate below 90 % (except from the One-Out technique). Since the One-Out validation method is quite prone to overtrain the implemented modules, the presented values are indicative for comparison purposes when all the available input set is exploited for training the algorithms. Especially, for the balanced 7-fold training and the generalized 3-fold validation the ANS supremacy was quite obvious, with corresponding performances of 95, 12 %and 92, 68 %, and therefore, favoring the development of generic classifiers based on ANS.

Table 7: Classification accuracy for listening tests according to sounds' FMC manifestation

	"Switch ON"	"Switch OFF"
1st Group	61, 26 %	51, 07 %
2nd Group	61, 57 %	45, 98 %
Average	61, 42 %	48, 52 %

The partial efficiency of class "Switch OFF" reached

Table 8: Machine learning tests results

ANS			
N. of Folds	$\mathbf{PR}_{overall}$	$\mathbf{PR}_{open}$	$\mathbf{PR}_{close}$
3-fold	92,68 %	92,59 %	92,86 %
7 fold	95,12 %	92,59 %	100 %
One-Out	97,56 %	96,30 %	100 %
Logistic Regression			
3 fold	85,37 %	85,19 %	85,71 %
7 fold	85,37 %	85,19 %	85,71 %
One-Out	90,24 %	92,59 %	85,71 %
SMO			
3 fold	90,24 %	88,89 %	92,86 %
7 fold	90,24 %	96,30 %	78,57 %
One-Out	95,12 %	96,30 %	92,86 %

100 % in ANS and the lowest value of 78,57 % with SMO algorithm, while Logistic Regression noted a steady rate of 85,71 %. The maximum and minimum partial discrimination rate of class "Switch ON" was 96,30 % for SMO algorithm (7-fold validation and One-Out technique) and 85,19 % for Logistic Regression respectively, while the corresponding value for ANS was 92,59 %. But when 3-fold validation was employed, in order to decrease the training iterations, ANS retained the best performance of 92,59 % for the class "Switch ON" compared to the decreased discrimination rate of 88,89 % of SMO algorithm. Consequently, in all cases, supervised training via ANS resulted in increased and more balanced overall and partial classification performances, promoting the construction of generic models for discriminating open/close sounds.

5. CONCLUSIONS AND FUTURE WORK

In the present work an evaluation of the intuitive conveyance of messages in HMI was performed. Moreover, the key technical features involved in this process were examined. To this aim, sounds commonly used in consumer devices for "Switch ON" and "Switch OFF" FMC manifestations were recorded, processed and analyzed. Conveyance's evaluation was accomplished using listening tests whereas the technical features involved were examined through extensive machine learning tests.

The high scores of the latter tests accuracy results indicate that there are underlying patterns used in developing sound motifs for FMC audio manifestation. Nevertheless, human participants exhibit an inability to correctly classify them, especially when it comes to the "Switch OFF" FMC. Despite this result, the small difference in the classification accuracy in the "Switch ON" FMC for both listeners' groups can indicate that both developers of sounds and listeners share a common perception for that particular FMC audio manifestation, not affected by listener's prior experience. This is an interesting result that may initiate future investigations for modeling this common sound design approaches followed by developers.

On the other hand, regarding the "Switch OFF" FMC category, the need for previous experience towards successful classification seem to emerge, by considering the accuracy difference of participants' groups (5,09 %). Nevertheless, the failure in correct classification from human listeners' indicate that further investigation is needed. What has to be noted is that the results of machine learning tests seem to denote that sound developers use common motifs for the "Switch OFF" FMC manifestation audio, allowing an accuracy score from ANS up to 100 %. However, these motifs seem not to be perceived correctly by listeners'.

Considering on one hand the not so wide data set and, on the other hand, the low accuracy values for the "Switch OFF" FMC audio manifestation, further investigation of FMC messages' conveyance through the audio channel seems to be needed. Additionally, taking into account the difficulties emerged in gathering the audio data set considered in this work, a joint research from manufacturers and researchers could lead to an extensive research potential for exploiting the phenomena and mechanisms involved in auditory HMI. The benefits from such a research could be the future development of common patterns for messages conveyance, capable to be accurately perceived by humans, within a broad HMI context. In turn, this could lead to a common audio communication code for HMI through the sound channel. Finally, the results of the present work can potentially lead to an extended research on the technical characteristics involved in intuitive and naturally occurred conveyance of auditory messages for HMI.

6. REFERENCES

[1] Walker B., and Kramer G., "Mappings and Metaphors in Auditory Displays: An Experimental Assessment," *ACM Transactions on Applied Perception (TAP)*, vol. 2, n. 4, pp. 407-412 (2005 Oct.).

[2] Gygi B., and Shafiro V., "From signal to substance and back: Insights from environmental sound research to auditory display design," presented in the 6th international conference on Auditory Display, Copenhagen, Denmark, May 18-22, 2009.

[3] Tzanetakis G., and Cook P., "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, n. 5, pp. 293-302 (2002 Jul.).

[4] Tzanetakis G., "Manipulation, analysis and retrieval systems for audio signals," PhD, Computer Science, Princeton University, Princeton, New Jersey, U.S.A., 2002.

[5] Lie L., et al., "Automatic mood detection and tracking of music audio signals," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, pp. 5-18 (2006 Jan).

[6] Drossos K., et al., "Emergency Voice/Stress - level Combined Recognition for Intelligent House Applications," presented in 132 Audio Engineering Society Convention, Budapest, Hungary, 2012.

[7] Chia-Hung Y., et al., "An efficient emotion detection scheme for popular music," presented at IEEE International Symposium on in Circuits and Systems 2009 (ISCAS 2009), 2009, pp. 1799-1802.

[8] Lu L., et al., "Content-based audio classification and segmentation by using support vector machines," *Multimedia Systems*, vol. 8, n. 6, pp. 482-492, (2003 Apr.).

[9] Drossos K., Floros A., Kanellopoulos G. N., "Affective acoustic ecology: towards emotionally enhanced sound events," presented at he 7th Audio Mostly Conference: A Conference on Interaction with Sound, pp. 109-116, Corfu-Greece, 2012.

[10] Roma G., et al., "Ecological acoustics perspective for content-based retrieval of environmental sounds," *EURASIP Journal on Audio, Speech, and Music Processing - Special issue on environmental sound synthesis, processing, and retrieval*, vol. 2010, n. 7, (2010 Jan.).

[11] Lartillot O., et al., "A Matlab Toolbox for Music Information Retrieval," in C. Preisach, H. Burkhardt, L. Schmidt- Thieme, R. Decker (Eds.), *Data Analysis, Machine Learning and Applications, Studies in Classification, Data Analysis, and Knowledge Organization*, Springer-Verlag, 2008.

[12] Hall M., et al., "The WEKA Data Mining Software: An Update," *SIGKDD Explorations*, vol. 11, n. 1, pp. 10-18 (2009).

[13] Kotsakis R., Kalliris G., and Dimoulas C., "Investigation of broadcast-audio semantic analysis scenarios employing radio-programme-adaptive pattern classification," *Speech Communication*, vol. 54, n. 6, pp. 743-762 (2012).

[14] Kotsakis R., Kalliris G., and Dimoulas C., "Investigation of salient audio-features for pattern-based semantic content analysis of radio productions," presented at the 132nd AES Convention, Budapest, 2012.

[15] Dimoulas C., Papanikolaou G., and Petridis V., "Pattern Classification and Audiovisual Content Management techniques using Hybrid Expert Systems: a video-assisted Bioacoustics Application in Abdominal Sounds Pattern Analysis," *Expert Systems with Applications*, vol. 38, n. 10, pp. 13082-13093 (2011).

[16] Vegiris C., Dimoulas C., and Papanikolaou G., "Audio content annotation, description and management using joint audio detection, segmentation and classification techniques," presented in 126th Audio Engineering Society Convention, Munich (2009).

[17] Hosmer D., and Lemeshow S., "Applied logistic regression (2nd ed.)," New York, Wiley, 2000.