Survey on Sound and Video Analysis Methods for Monitoring Face-to-Face Module Delivery

https://doi.org/10.3991/ijet.v14i08.9813

Andreas E. Papadakis (✉), Eleni Tsalera
School of Pedagogical and Technological Education, Athens, Greece
apapadakis@aspete.gr

Maria Samarakou
University of West Attica, Athens, Greece

Abstract—The objective of this work is to identify unobtrusive methodologies that allow the monitoring and understanding of the educational environment, during face-to-face activities, through capturing and processing of sound and video signals. It is a survey on applications and techniques that exploit these two signals (sound and video) retrieved in classrooms, offices and other spaces. We categorize such applications based upon the high level characteristics extracted from the analysis of the low level features of the sound and video signals. Through the overview of these technologies, we attempt to achieve a degree of understanding of the human behavior in a smart classroom, on behalf of the students and the teacher. Additionally, we illustrate open-research points for further investigation.

Keywords—Context extraction, sound and video analysis, smart spaces

1 Introduction

While remote and virtual learning environments are being massively investigated, explored and extended, based on current technological capabilities, the face-to-face learning scenario has remained, largely, unchanged. Students who actively engage in the module delivery tend to understand, learn, remember more, and be more able to appreciate the relevance of what they have learned, than students who passively receive.

In the face-to-face learning scenarios, the module delivery is performed within classrooms that students physically attend. In the classroom, a rich set of activities take place and a wealth of signals are generated by the students and the teachers. Such signals and data can offer important information related to student engagement and emotions (which can be predictors of learning). Student attention, collaboration as well as the characteristics of module delivery and teaching styles (e.g. monologues or based on discussion, dialogues and experimentation) are well hidden in these signals. This information can support enriched learning analytics and allow the stakeholders to take informed education – related decisions.

While these signals offer live feedback to the teacher and the instructor, who can make adaptations and corrective movements during the module delivery, they remain largely unexploited in terms of their machine-based capturing, processing, analysis, understanding and exploitations.

This is the scope of this paper as we perform a survey of research activities that are related to the monitoring, processing and understating of signals in an unobtrusive way in smart spaces. These can be already related to education or they can be potentially exploited in learning scenarios. To support the unobtrusive characteristics of the monitoring framework we focus on sound and video signals.

The structure of the paper is the following: Section 2 describes the framework of our survey and sets the criteria for our decision, including the signal and technology selections and the typical architectural layering. Section 3 analyzes the sound signals and the features that are extracted in the research activities that we have considered. We also correlate this analysis with behavioural features that are of interest during module delivery. Section 4 examines the features extracted by video signals and performs a similar correlation. Finally, Section 5 presents the conclusions and suggests some open points for future investigation and research.

2 Research Framework

The personalized diagnosis, assistance and evaluation of students in open learning environments are challenging tasks, especially when considering real-time, classroom conditions. In the past the members of our team have investigated and designed an open learning environment to monitor the comprehension of students, assess their prior knowledge, build individual learner profiles, provide personalized assistance and, finally, evaluate their performance by using artificial intelligence [1]. Monitoring a student's behavior in text comprehension activities allows the inference of the student cognitive profile and selection of the personalized feedback and support [2].

Our focus is the educational environment itself, and specifically within the face-to-face module delivery. We are investigating research efforts, possibly focused upon different application domains that can be adapted and applied for the module delivery.

In principle we consider the concept of the Smart Environment (SE) defined as able to acquire and apply knowledge about this environment and its inhabitants in order to improve their experience in that environment [3]. This can be achieved with information acquisition from intelligent sensor devices, communications among sensors, enhanced services offered by intelligent devices and predictive and decision-making capabilities.

2.1 Signal and technology selection framework

The infrastructures and the systems under examination have to be unobtrusive in terms of the education processes both for the students and the teacher / instructor. As already discussed, to support this requirement we select methodologies and technologies monitoring and analyzing sound and video signals through typical microphones and cameras, i.e. without special equipment held by the students. The objective of our survey is to identify technologies and mechanisms able to monitor the classroom conditions through the generated signals during module delivery. In the research efforts that have been surveyed, the monitored space is indoor and preferably, but not necessary, related to education / classroom. The infrastructure should be preferable based on cost effective tools and easily deployed tools and equipment.

In terms of the architectural approach, we consider two layers, consisting of an operational and an intelligent layers allowing for interaction with the monitored space and themselves, as sufficiently flexible and scalable to accommodate the modules we consider [4].

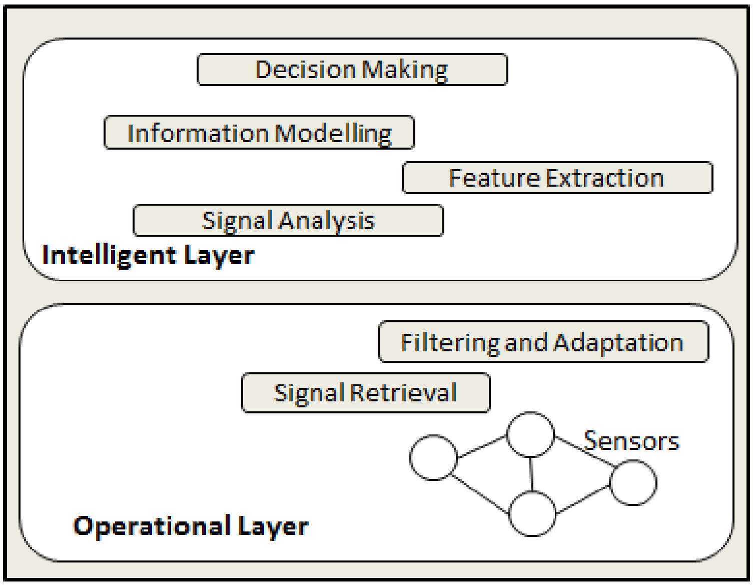

Fig. 1. Typical architecture consisting of operation and intelligent layers

The operational layer is composed by the sensing (and possibly actuating) capabilities mainly including the sensors and the signal retrieval and adaptation modules. The intelligent layer includes the signal processing and interpretation techniques and allows understanding of the monitored signals, the environment's state and information representation.

3 Sound Signals

Acoustic signals, sound, can provide useful information related to a smart space and especially for the face-to-face scenario, the classroom conditions and participation of the teacher and the students. Aspects of interest include the number of speakers in the room, at any given time, considering two extremes: the teacher's monologue to the generalized dialogue between the students along with multiple intermediate states of students contributing to the module delivery with observations, questions and organized and controlled dialogues.

In the following, we group the capabilities of sound processing in smart environments that can be associated with the needs of a monitored classroom, in an escalating complexity, and refer to the identified research efforts.

Sound Identification: In [5] sound analysis is performed through structured statistical modeling. The Smart Ambient Sound Analyzer platform supports different ambient audio mining tasks (including audio classification and location estimation) extracting acoustic features from sound components (e.g., music, voice and background), and translates them into structured information. The system integrates mixture models and an SVM algorithm into a unified classification framework.

Sound Detection is performed in [6], where the designed system allows for detecting sound events from input streams. The Sound/Speech discrimination functionality allows for discriminating speech from other sounds to extract voice commands, while the classification recognizes daily living sounds and the speech recognition applies speech recognition to events classified as speech.

Voice Activity Detection: Voice activity detection allows classification of input signal frames based on feature estimation in two classes: speech activity and non-speech events (pauses, silence, or background noise). This separation has been performed in another application domain (related to movies) in [7].

Speech Loudness and Clarity characterization: The environment noise level in classrooms has been correlated with the intensity and quality of the teachers' voice [8] using the Pearson Correlation Test. The mean of sound pressure level is calculated in specific locations in the classrooms. Voices were classified according to the GRABSI parameters including grade, roughness, asthenia, breathiness, strain and instability.

Sound Localization: More elaborate sound analysis scenarios involve the localization of the sound (i.e. to locate sound sources in space) so that the sound generation context can be understood in a more meaningful way and the identification of sound sources can be confronted. The human binaural auditory system is capable of analysis of auditory scenes and sound localization, in a classroom environment it is a challenging task to localize the student that posed a question without Line of Sight (e.g. if the teacher is writing on the board). The mechanism of localization is well explained in physiology in terms of interaural difference cues. The listener's head interrupt the sound path from the source to the far ear, resulting in a difference of pressure level and time of arrival (or phase) between the two ears, with neurons being able to measure such differences [9].

According to [10], the existing Sound Source Localization technologies can be categorized into three groups: Time delay estimation (TDE), beamforming method and machine learning methods. TDE is enhanced using multichannel cross-correlation-coefficient algorithm to exploit spatial and temporal information among multiple microphones in [11]. The measurement of the active sound sources and the estimation of their corresponding DOAs is achieved in [12] by representing the observed signals in Time-Frequency zones where each source is dominant from the others. In [13] sound localization is performed based upon spatially distributed sensors calculating Time Differences of Arrival (TDOAs) between microphones of the same sensor. In [14] also sound localization is realized using TDE but the novelty here is the usage of arbitrarily shaped non-coplanar microphone arrays.

In [15] a microphone array sensor has also been deployed, so that through the analysis of the signals to detect the number of speakers within a moving window-based time interval. The system computes the TDOA (Time Difference of Arrival) between each microphone pair by using the received acoustic signal patterns and identifying the time lag between each signal pair. In order to confront the room reverberations and background noises, the system uses generalized cross correlation with phase transform (GCC-PHAT). The system locates the direction of the speakers, and uses this information to detect the number of speakers in the target environment.

The beamforming method follows subspace approaches (exploiting the orthogonality between signal and noise subspaces) and beamscan approaches localize the array signals into one direction (such as the Steered Response Power Phase Transform) [16], [17], [18], [19]. Machine learning approaches are supervised learning methods, including support vector machine [13], multilayer perceptron neural network [20] and Gaussian mixture model [21].

Physical Room Acoustic Feature Modeling: Given the challenges in estimating the accuracy of direction of arrival (DOA) in indoor environments due to echo, multipath propagation and spectral distortions as well as due to the fact that noise sources may have similar spectral characteristics with the signal monitoring, the study and modeling of the space acoustic features may improve the DOA estimation. The acoustic features of the physical rooms can be pre-studied before localization. This kind of data driven training methods can be more effective especially when the environment is too complex to be modeled.

Parameters to identify – Situations to understand: In the following table, we correlate the aforementioned sound analysis results with parameters related to the classroom activities.

Table 1. Sound-related features and corresponding situations

Sound processing	Corresponding situations
Sound detection	Presence of people Verification of module delivery
Sound identification and voice detection	People speaking (teacher or students) Frequency of speaking and temporal coverage of speaking Number and types of other sounds being heard in the classroom Quality of the speech Noise in the environment
Sound localization	Number of people speaking Type of module delivery (teacher monologue, discussion) Interest on behalf of the participants

The main challenges in such environments include the existence of multiple people (teacher and students) that can be potential sound sources with different vocal characteristics, the echoing and the reverberation.

4 Video Signals

Video capturing and processing can unobtrusively retrieve indications produced by the students and the teachers during the module delivery. Such features may span from presence to activity and expression recognition. In the following we group the capabilities of video processing in smart environments that can be associated with the needs of a monitored classroom.

4.1 Human face detection and identification

Human detection and identification can be some of the first objectives of image capturing and analysis. In [22] and [23] a rotating camera is positioned centrally in the front of the classroom and captures frontal images from the students. Histogram normalization allows for contrast enhancement in the spatial domain and median filtering allows removal of noise. Skin classification allows separating pixels related with skin (making them white). For face detection, Haar classifiers are used. The detected faces are compared with the database and when a face is recognized the attendance is marked on the server. In [24], one camera retrieves the occupied seats and another captures the images of student's face.

4.2 Facial expression recognition

Recognition of the students' emotions belongs to the objectives of [25], [26] which aim at recognizing facial expressions and head gestures. Based on Ekman's analyses [27] and using a facial expression recognition system [28] it is attempted to infer mental states from a video stream of facial events in real-time. The system, capturing at 30 fps, locates and tracks 24 feature points on the face and uses motion, shape and color deformations to identify facial and head movements (e.g. head pitch, lip corner pull) and communicative gestures (e.g. head nod, smile). The work in [29] detects facial expressions connected with frustration. The red-eye effect is used to track pupils and Hidden Markov Models (HMMs) detect head nods, shakes and blinks. Support vector machine (SVM) is used to compute the two opposite expressions of smiles or fidgets.

4.3 Gaze recognition

In [30] and [31] a system of cameras are directed towards the students' audience. Using Pointing'04 database [32] faces are detected using a probabilistic detection of skin chrominance by normalizing the red and green components of the RGB color vector by the intensity (R+G+B). Face position and thus gaze destination are estimated using a zeroth order Kalman Filter which permits the process to be focused on the face region. Head orientation and gaze estimation up to the distance of 4 meters away from the black-board was effectively captured by re-training a face detector and developing additional testing data-sets. A mobile eye-tracker is also needed (possibly worn by the teacher) which may lead to distraction for both the teacher and the students.

4.4 Posture and gesture recognition

Human posture and gesture perception supports the understanding of human behavior and attitude. These features are extracted in [33] during human-machine interaction. To achieve real-time response, the upper body part is analyzed as it projects information about many human activities. The body pose estimation is analyzed in two parts: first skin color is used to track head and hand blobs and then a silhouette is shaped from the lengths of the shoulders and the neck. The 3-D movements of head and hands are tracked using multi-view input. This approach presupposes that the person stands in front of the camera with straight arms and facing forwards at the beginning of each session in order to construct a body model for initialization.

An investigation of building spatio-temporal models of human actions that can support categorization and recognition of simple action classes is conducted in [34]. Action recognition is realized in two stages: the characteristics of motion are extracted from visual input and then these characteristics are classified into action classes. This approach considers variations in viewpoints around the central axis of human body and proposes a representation based on Fourier analysis of motion history volumes in cylindrical coordinates. 3-D motion descriptors have been extracted that support meaningful categorization of simple action classes.

Fusing of facial expression and body gesture is suggested in [35], considering that many of the spatial-temporal features detected are similar. Combining visual channels of facial expression and body gesture is a potential way to accomplish effective affect analysis [36]. Two cameras are used for capturing the facial expression and the body movements. The Canonical Correlation Analysis (CCA) is used in order to find pairs of base vectors (i.e. canonical factors) for two variables such that the correlations between the projections of variables onto these canonical factors are mutually maximized.

4.5 Human activity recognition

In [30], [31] a camera, placed at the back of the classroom, captures teacher's actions and slight changes in the scene. Major body movement and gesticulation are recognized due to the long distance. The tracker/detector [37] used to track teacher's motion in 1D horizontal location in front of the blackboard. The tracker estimates the object's motion between consecutive frames and the detector treats every frame as independent and performs full scanning of the image to localize all appearances that have been observed and learned in the past in order to generate training examples and thus to avoid same errors in the future.

4.6 Modeling of environment

In [38] human action is recognized from video input in an environment for which prior knowledge is available and the possible actions are pre-categorized (e.g. entering/exiting the scene, standing up, sitting down, using a computer terminal). Before detecting activities, statistical information about monitored area is computed. This reduces the effect of lighting conditions and other factors that change over time. The prior knowledge of the area and the limitation of under consideration actions simplify the actions' determination.

Table 2. Video-related features and corresponding situations

Video processing	Corresponding situations
Movement detection	Presence of people Verification of module delivery Level of attention and quitness
Face detection	Counting of people Identification of attendants
Facial expression identification Gaze detection	Level of attention and interest Interactivity among participants
Gesture and activity recognition	Level of attention and discipline Phases of module delivery

The main challenges in such environments include the existence of multiple people (teacher and students) that can be potential sources of movements. Furthermore the complexity of actions, the proximity of people and the difficulty to detect a person if there is no clear view. These challenges are partially confronted with multiple cameras (multi-view), increased resolution and the fusion – based enhancements.

5 Conclusion and Future Work

In this paper, we have surveyed applications based on audio and video analysis able to detect and recognize human behaviours and subsequently to predict and enhance the imminent procedures and activities. As discussed we have focused on the education environment for face-to-face module delivery but the research efforts identified come from other applications domains (including the smart spaces). Our focus has been on sound and video and the main restriction has been the unobtrusive nature. For both areas a set of methodologies have been identified, while the achieved results have been correlated with behaviours and features met in the classroom.

For audio analyses, we have identified research works and methodologies, which recognize speech from other sounds; detect dialogue; recognize the number of speakers; estimates the level of noise in a room; analyzes the intensity and clarity of the voice and localize the sound source. For video analyses, the identified approaches concentrate on the extraction of diverse human activities from human presence to facial expression recognition, posture and gesture recognition and activity recognition.

The maturity and the results presented in the sources identified allow confidence that this field can be further developed through the consolidation of such methods. On the other hand, we recognize that the classroom environment is highly challenging, due to the co-existence of multiple people acting as signal (sound and movement) sources. Furthermore behaviour is not strictly formulated in such environments and unexpected (or undefined) activities may take place. In this view, systems should recognize concurrent activities from a person rather than focusing on a single activity each time.

6 References

[1] Papadakis A., Fylladitakis E. D., Hatziapostolou A., Tsaganou G. & Früh W. G. Samarakou M., "An open learning environment for the diagnosis, assistance and evaluation of students based on artificial intelligence.," vol. 9, no. 3, pp. 36-44, 2014.

[2] Grammatiki Tsaganou, Andreas Papadakis, John Gelegenis, Emmanouil D Fylladitakis, Maria Grigoriadou Maria Samarakou, "Monitoring the text comprehension of students for profiling in ReTuDiS," *Journal of Information Technology and Application in Education (JITAE)*, Dec. 2013.

[3] D.J. & Das, S.K. Cook, "How smart are our environments? An updated look at the state of the art. ," *Pervasive and Mobile Computing vol. 3*, 2007.

[4] C. F., Barroso, J., & Ramos, C. Freitas, "A Survey on Smart Meeting Rooms and Open Issues," *International Journal of Smart Home*, vol. 9, no. 9, pp. 13-20, 2015. https://doi.org/10.14257/ijsh.2015.9.9.02

[5] J., Nie, L., & Chua, T. S. Shen, "Smart ambient sound analysis via structured statistical modeling," in *International Conference on Multimedia Modeling*, 2016, pp. 231-243.

[6] M. A., Lecouteux, B., Vacher, M., Portet, F., Istrate, D., Dorizzi, B., & Boudy, J. Sehili, "Sound Environment Analysis in Smart Home.," in *AMI 2012, http://sweet-home.imag.fr/documents/2012_AMI_Sehili.pdf*, Berlin, Heidelberg, 2012, pp. 208-223.

[7] M., Ververidis, D., Panagakis, Y., Kotropoulos, C., Maragos, P., & Pitas, I. Kotti, "Audio-Assisted Movie Dialogue Detection," *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 18, NO. 11*, vol. 18, no. 11, pp. 1618-1627, November 2008.

[8] R. F., Bertoncello, F., Zanchetta, S., & Dragone, M. L. S. Guidini, "Correlations between classroom environmental noise and teachers' voice," *Revista da Sociedade Brasileira de Fonoaudiologia (vol.17 no.4)*, vol. 14, no. 4, pp. 398-404, December 2012.

[9] D. Wang and G.J. Brown, "Computational auditory scene analysis: Principles, algorithms and applications," *Wiley-IEEE press.*, 2006.

[10] Y., Chen, J., Yuen, C., & Rahardja, S. Sun, "Indoor Sound Source Localization with Probabilistic Neural Network," *IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS*, vol. 65, no. 8, pp. 6403-6413, December 2017.

[11] H., Wu, L., Lu, J., Qiu, X., & Chen, J. He, "Time difference of arrival estimation exploiting multichannel spatio-temporal prediction.," vol. 21(3), pp. 463-475, 2013.

[12] D., Griffin, A., Puigt, M., & Mouchtaris, A. Pavlidi, "Real-time multiple sound source localization and counting using a circular microphone array.," vol. 21(10), pp. 2193-2206, 2013.

[13] H., & Ser, W. Chen, "Acoustic source localization using LS-SVMs without calibration of microphone arrays.," no. Circuits and Systems, ISCAS, pp. 1863-1866, May 2009.

[14] X., & Horaud, R. Alameda-Pineda, "A geometric approach to sound source localization from time-delay estimates.," vol. 22(6), pp. 1082-1095, 2014.

[15] Jongjun Park, Hyunhak Kim, Jongarm Jun, Sang Hyuk Son, Taejoon Park and JeongGil Ko Homin Park, "ReLiSCE: Utilizing Resource-Limited Sensors for Office Activity Context Extraction," *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS (Vol. 45, No. 8)*, August 2015.

[16] J., Martín-Arguedas, C. J., Macias-Guarasa, J., Pizarro, D., & Mazo, M. Velasco, "Proposal and validation of an analytical generative model of SRP-PHAT power maps in reverberant scenarios.," vol. 119, pp. 209-228, 2016.

[17] B., & Aarabi, P. Mungamuru, "Enhanced sound localization.," vol. 34(3), pp. 1526-1540, 2004.

[18] J. P., Benesty, J., & Affes, S. Dmochowski, "A generalized steered response power method for computationally viable source localization. ," vol. 15(8), pp. 2510-2526, 2007.

[19] D., Lee, T., & Cho, Y. Yook, "Fast sound source localization using two-level search space clustering. ," vol. 46(1), pp. 20-26, 2016.

[20] X., Zhao, S., Zhong, X., Jones, D. L., Chng, E. S., & Li, H. Xiao, "A learning-based approach to direction of arrival estimation in noisy and reverberant environments.," in *IEEE International Conference on IEEE.*, 2015, pp. 2814-2818.

[21] X., & Liu, H. Li, "Sound source localization for HRI using FOC-based time difference feature and spatial grid matching. ," vol. 43(4), pp. 1199-1212, 2013.

[22] Yousaf M., Ahmad W., Baig M.I. Balcoh N., "Algorithm for efficient attendance management: Face recognition based approach," vol. 9.4:146, 2012.

[23] Nouman H.M.F., Mushtaq M.O., Raza B., Tayyab A., Talib M.W. Fuzail M., "Face detection for attendance of class' students.," vol. 5(4), 2014.

[24] Shoji T., Weijane L.I.N., Kakusho K., Minoh M. Kawaguchi Y., "Face recognition-based lecture attendance system," 2005.

[25] Cooper D., Burlescon W., Woolf B. P., Muldner K., Christopherson R. Arroyo I., "Emotion sensor go to school," *AIED*, pp. Vol, 200, p.17-24, 2009.

[26] Arroyo I., Woolf B., Burlescon W., El Kaliouby R., Eydgahi H. Dragon T., "Viewing student affect and learning through classroom observation and physical sensors," in *International Conference on Intelligent Tutoring Systems*, Berlin, Heidelberg, 2008, pp. 29-30.

[27] Ekman P., "Facial expressions of emotion: New fndings, new questions," *Psychological Science*, pp. 34-38, 1992. https://doi.org/10.1111/j.1467-9280.1992.tb00253.x

[28] Robinson P. El Kaliouby R., "Real-time inference of complex mental states from facial expressions and head gestures," *Real-time vision for human-computer interaction, Springer, Boston, MA*, pp. 181-200, 2005.

[29] Burlescon W., Picard R.W. Kapoor A., "Automatic prediction of frustration.," vol. 65(8), pp. 724-736, 2007.

[30] Dillenbourg P. Raca M., "Holistic analysis of the classroom," in Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge, 2014, pp. 13-20.

[31] Dillenbourg P. Raca M., "System for assessing classroom attention," in *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 2013, pp. 265-269.

[32] Hall D., Crowley J.L. Gourier N., "Estimating face orientation from robust detection of salient facial features," , 2004, p. ICPR International Workshop on Visual Observation of Deictic Gestures.

[33] Trivedi M.M. Tran C., "3-D posture and gesture recognition for interactivity in smart spaces," *IEEE Transactions on Industrial Informatics*, vol. 8(1), pp. 178-187, 2012. https://doi.org/10.1109/TII.2011.2172450

[34] Ronfard R., Boyer E. Weinland D., "Free viewpoint action recognition using motion history volumes," vol. 104(2-3), pp. 249-257, 2006.

[35] C., Gong, S., & McOwan, P. W. Shan, "Beyond Facial Expressions: Learning Human Emotion from Body Gestures.," , 2007, pp. 1-10.

[36] N., & Rosenthal, R. Ambady, "Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis.," vol. 111, no. 2, pp. 256-274, 1992.

[37] Mikolajczyk K., Mata J. Kala Z., "Tracking-learning-detection," *IEEE transactions on pattern analysis and machine intelligence*, pp. 34.7:1409, 2012.

[38] Shah M. Ayers D., "Monitoring human behavior from video taken in an office environment," vol. 19(12), pp. 833-846, 2001.

[39] J. Shen et al., "Smart ambient sound analysis via structured statistical modeling," in *International Conference on Multimedia Modeling*, Miami , 2016. https://doi.org/10.1007/978-3-319-27674-8_21

[40] Delandshere G., Danish J.A. Andrade A., "Using Multimodal Learning Analytics to Model Student Behavior: A Systematic Analysis of Epistemological Framing," *Journal Learning Analytics*, pp. 282-306, 2016.

[41] Wagealla W., English C., Terzis S. Nixon P., "Security, Privacy and Trust Issues in Smart Environmnents.," 2005.

[42] B., Talmon, R., & Gannot, S. Laufer-Goldshtein, "Semi-supervised sound source localization based on manifold regularization.," vol. 24(8), pp. 1393-1407, 2016.

[43] A., Antonacci, F., Sarti, A., & Tubaro, S. Canclini, "Acoustic source localization with distributed asynchronous microphone networks.," vol. 21(2), pp. 439-443, 2013.

[44] D.J. & Das, S.K. Cook, *Smart Environments: Technology, Protocols, Applications.*: John Wiley & Sons., 2004, vol. 43.

[45] B. P. L., Wijerathne, N., Ng, B. K. K., & Yuen, C. Lau, "Sensor Fusion for Public Space Utilization Monitoring in a Smart City," *IEEE Internet of Things Journal*, vol. 5(2), pp. 473-481, April 2018. https://doi.org/10.1109/JIOT.2017.2748987

[46] H. Hotelling, "Relations between two sets of variates.," vol. 28(3/4), pp. 321-377, 1936.

[47] S., Papadopoulou, E., Taylor, N. K., & Williams, M. H. Gallacher, "Learning user preferences for adaptive pervasive environments: An incremental and temporal approach.," vol. 8, no. 1, 2013.

7 Authors

Andreas Papadakis (Dr-Ing.) holds degree and PhD from the Department of Electrical and Computing Engineering in NTUA (National Technical University f Athens). He is an Associate Professor in the department of Electrical and Electronics Engineering Educators in the School of Pedagogical and Technological Education (ASPETE), Athens, Greece. He has published more than 50 papers in refereed scientific journal and proceedings of International conferences and he has participated in more than 10 RTD European projects in the area of innovative Internet and telecommunications services. His research work has more than 200 citations and he is a regular reviewer of international journals.

Eleni Tsalera is Electronics Engineer with Master Diploma in Electronics and Radioelectrology; she is currently Ph.D. candidate in University of West Attica at the Department of Informatics and Computer Engineering. She also works as research associate in School of Pedagogical and Technological Education (ASPETE) in Athens (Greece) since 2009.

Maria Samarakou (Dr-Ing.) holds degree and PhD from the Department of Physics in National and Kapodistrian University of Athens. She is a Professor at the department of Informatics and Computer Engineering, University of West Attica and Director of Laboratory of Educational Technology and e-Learning Systems. Her research work has contributed to the design of educational environments, intelligent tutoring systems, artificial intelligence, energy management, web based education and computer science education. She has undertaken more than 20 National and European projects in research and technology development as coordinator/project manager or main researcher. She has published more than 100 papers in refereed scientific journal and proceedings of International and National congresses on topics in the field of simulation, optimization, expert systems, artificial intelligence and educational technology. She has more than 400 citations in scientific articles and she is Reviewer in various international scientific journals and conferences.

Article submitted 2018-11-03. Resubmitted 2018-12-13. Final acceptance 2018-12-15. Final version published as submitted by the authors.