7. Abstracts of the publications grouped in categories “B4”, “G7” and “G8”

7.1 Abstracts of the publications grouped by criterion ”B.4”

[B4.1]. V. Markova, **T. Ganchev,** K. Kalinkov (2018). “Detection of Negative Emotions and High-Arousal Negative-Valence States on the Move”, *Proc. of the 4th International Scientific Conference on Advances in Wireless and Optical Communications,* RTUWO-2018, Riga, Latvia, November 15-16, 2018, pp. 61-65.

In paper [B4.1], we present the overall design and implementation of a wearable system that is capable to continuously monitor and register negative emotions, high levels of emotion arousal, and high-arousal-negative-valence states from physiological signals, such as skin conductivity and ECG. This system builds on the client-server architecture. The commercially available wireless data acquisition devices Shimmer3 GSR+ and Shimmer3 ECG are used for the acquisition of physiological signals, which are then transmitted over a Bluetooth channel to a mobile phone. The mobile phone hosts the user interface and implements the data aggregation and transmission to the server, which carries all signal processing and classification tasks. Purposely developed software implements all data processing and recognition tasks on the server side and the user interface and the data communication on the client side. We report evaluation results for various setups of the binary detectors of negative emotions, high level of emotional arousal, and high-arousal negative-valence states.

[B4.2]. V. Markova, **T. Ganchev** (2018). "Constrained Attribute Selection for Stress Detection Based on Physiological Signals," *Proc. of the 2018 International Conference on Sensors, Signal and Image Processing,* SSIP-2018, Prague, Czech, October 12-14, 2018. pp. 41-45.

In [B4.2], we present a constrained attribute selection method that makes use of feature assessment based on the Fisher's separation criterion followed by variety reduction post-processing. The post-processing incorporates task-specific constrain into the feature selection process, as this is expected to facilitate the subsequent data modeling and classification stages.

Here we validate the proposed method in an experimental setup oriented towards acute stress detection based on physiological signals. For this purpose, we made use of a subset of the ASCERTAIN database, which contains physiological records (ECG signals and surface resistance of the skin) for 58 people. Based on these physiological signals are calculated 43 descriptors, from which through a purposely designed selection procedure, are selected the 18 most relevant to the problem, i.e. these are selected based on prior knowledge of the problem.

The experimental results support that the proposed method brings advantage, when compared to three other cases: (i) the full set of features, (ii) a subset selected based on prior knowledge, (iii) and a subset selected based solely on Fisher's separation criterion.

Compared to these baseline methods, a relative average improvement in the recognition accuracy of acute stress conditions was achieved, with an average (for all 58 subjects) improvement of 5.4%.

[B4.3]. V. Markova, **T. Ganchev** (2018). "Three-step Attribute Selection for Stress Detection Based on Physiological Signals," *Proc. XXVII International Scientific Conference Electronics - ET2018*, Sozopol, Bulgaria, September 13–15, 2018. Article number 8549658

We present a three-step method for attribute selection that builds on person-independent and person-specific feature assessment stages. The first two steps aim to select a person-independent subset of attributes that are repeatedly selected for a large population of users. Next, this selection is intersect with a person-specific subset derived from the Fisher's separation criterion. As a result, we obtain a subset of attributes, which is both task-specific and customized to the quality of data of each particular user. The proposed method was validated on the ASCERTAIN database in an experimental setup oriented towards high-arousal negative-valence detection based on physiological signals.
For that purpose we used ECG signal recordings and skin surface resistance for all 58 subjects. The experimental setup is directed to the recognition of physiological conditions, associated with stress, which are characterized by high levels of emotional arousal in negative emotions. A comparative experimental study of the classification accuracy was performed for eight sets of descriptors, including

(i) three different settings for the proposed method,
(ii) three different settings for the method of selection of the most frequently selected descriptors after Fisher selection,
(iii) the individual subsets of descriptors that are selected by Fisher method individually for each person,
(iv) the complete set of 43 descriptors.

The experimental results support that the proposed method offers advantage in terms of detection accuracy when compared to other subset selection strategies. As a result of the implementation of the new feature selection method, an improvement in the relative recognition accuracy of over 5.3% has been reported over the complete set of descriptors.

[B4.4]. V. Markova, **T. Ganchev** (2018). "Automated Recognition of Affect and Stress Evoked by Audio-Visual Stimuli," *Proc. of the Seventh Balkan Conference on Lighting*, BALKANLIGHT-2018, Varna, Bulgaria, September 20-22, 2018. Article number 8546887

In the present work, we consider the automated detection of negative emotions and high-arousal negative-valence (HANV) states, which are akin to acute stress occurring in specific context. We investigate the influence and intricacy, which subjective perception of negative emotions and acute stress brings to the process of automated recognition of these states. For that purpose, we experimentally evaluate the advantages of modelling based on the person-independent tags of the audio-video stimuli, and models built with the person- specific self-reported tags.

Regardless of the ultimate aim of detection, the two tagging approaches use alternative formulations of the task -- through the use of tags, which are:

1) person-independent tags -- to some extent objective, since are common to all people and were defined solely by the content of each specific stimulus (more or less consensual for a large group of people), or
2) person- specific self-reported tags -- subjective and individual for each potential user of this technology; They are derived from each individual's self-determination of the effect of the application of a particular stimulus.

Using these two formulations of the task, we explore the possibilities of creating automatic detectors of HANV states and of negative emotional states. The implications and complexity of the task in both formulations and the possibilities for automated recognition of these conditions are investigated. For this purpose, an experimental evaluation of the benefits of independent tag-based modeling was carried out, using the labels attached to the audio-video incentives by the database creators, as well as those obtained from the self-assessment of specific people. Based on the self-reported tags, which are obtained with only tiny extra effort, we report a relative improvement of the HANV detection accuracy with up to 5%.

[B4.5]. T. Ganchev, V. Markova, I. Lefterov, Y. Kalinin (2017). "Overall Design of the SLADE Data Acquisition System", Proc. of the IITI-2017, September 14-16, 2017, Varna, Bulgaria. *Advances in Intelligent Systems and Computing,* ISSN: 2194-5357, vol. 680, pp.56-65.

In [B4.5], we present the overall design of a data acquisition system developed for the needs of the SLADE (Stress Level and Emotional State Assessment Database) database. The database consists of synchronized EEG, ECG, skin temperature (ST), and galvanic skin response (GSR) recordings, used in stress level assessment and recognition of emotional states. SLADE will facilitate the development of automated tools and services for stress-level assessment and monitoring.

In addition to description of the quantitative and qualitative parameters of the physiological signals database, experimental work has been performed to validate the usefulness of the SLADE database. The paper presents experimental results of a baseline system for the recognition of (i) emotional arousal, (ii) negative emotional states, and (iii) conditions, characterized by a high degree of emotional arousal in negative emotional states. The latter are associated with the group states, having a relationship to acute stress. This report is described in detail the architecture of the system for recognition of emotional states, extraction of descriptors, post-processing of descriptors and the classification process.

The research presented in this paper is partially supported by the project НП7/2017 "Methods for gathering biomedical information through mobile information systems" funded by the Technical University of Varna.

[B4.6]. F. Feradov, I. Mporas, **T. Ganchev** (2017). "Evaluation of Cepstral Coefficients as Features in EEG-Based Emotional States Recognition," Proc. of the 2nd International Scientific Conference "Intelligent Information Technologies for Industry", IITI-2017, September 14-16, 2017, Varna, Bulgaria. *Advances in Intelligent Systems and Computing*, ISSN: 2194-5357, vol. 680, pp. 504-511

The study of physiological signals and the evaluation of their features are of great importance for the automated emotion detection, as these are directly connected with the successful modelling and classification of the states of interest. In [B4.6], we present an evaluation of the appropriateness of LFCC and the logarithmic energy of signals as features for automated recognition of negative emotional states in terms of recognition accuracy.

In particular, three sets of features are compared – features computed after frame-level segmentation of the signal; features computed after averaging of frame level descriptors; and features extracted from an entire EEG recording.

Comparative analysis of the various configurations of these three methods was carried out by an experimental protocol, using EEG recordings from the DEAP database of physiological signals. Participants' data are divided into two groups, depending on the evaluation of the liking criteria set by the participants. Records received score lower than 4, is flagged as "negative", while recording with a score higher than 4, were labeled as "other" (neutral or positive). The tests use data from 24 participants who, after the distribution of records, each of the two classes contains at least 20% of the total volume of EEG data. The usefulness of the different sets of descriptors was evaluated in the WEKA environment. The performance of the extracted features is evaluated using C4.5 classifier for 10, 15, 20, 30, 45, and 60 filters.

The resultant average classification accuracy varies within 70.7 % - 86.9%. The analysis of results shows that the accuracy of linear cepstral coefficients and spectral logarithmic energy is very close in cases where EEG signals are segmented by a window function, in which cases the highest achieved accuracy is observed. Both categories of descriptors when applying averaging EEG sections, are relatively low scores.

Similar results were observed with the use of descriptors, calculated from EEG non-segmented signal, by banks with a small number of filters -- 10 and 15 filter. When increasing the number of filters, however, an improvement of the accuracy of classification is observed. This dependence is more pronounced when using spectral logarithmic energy where the difference in the average accuracy of the classification between the use of 10 and of 60 filter 12.2%.

[B4.7]. Feradov F., **Ganchev T**. (2016). "Spatio-temporal EEG signal descriptors for recognition of negative emotional states," *Proc. of the International Scientific Conference Electronics,* ET-2016, September 12-14, 2016. IEEE eXplore.

In [B4.7], we present a study on novel signal descriptors for the purposes of automated recognition of negative emotional states from EEG signals – namely, the decorrelated values of the energy of the spatio-temporal distribution of EEG activity. We investigated the spatial relationships between EEG signals and created descriptors, which take into account such spatial correlations. We propose a new descriptors of the EEG signal, representing the *decorated energy of the time-space distribution of EEG activity* and evaluated its applicability for the recognition of negative emotional states.

The feature computation process consists of four steps:

1) conversion of the original multichannel EEG signal into a time-space representation of the signal. The transformation is accomplished by applying a short-time Fourier transform to the spatial dimension of the signal simultaneously across all channels.
2) the EEG signal spectrogram is scaled by a set of overlapping triangular window functions.
3) the logarithmic energy of the spatial activity for the selected period of the EEG signal is calculated. Taking the calculated space-time energy for each window, we get a representation of the spectrogram scaled by time.
4) a discrete cosine transform is applied. Thus the energy calculated for each time window, is decorrelated in order to obtain the proposed Decorrelated Spatio-Temporal Coefficients (DSTC).

An experimental evaluation on the detection of negative emotional states is presented. Quantitative assessment, in terms of classification accuracy, of the usefulness of descriptors is performed in an experimental protocol using EEG signals from the DEAP database and grouping of recordings according to two different criteria - "liking" and "valence".

Using the extracted features person-specific SVM models are created. The classification accuracy of the proposed features is evaluated using two experimental setups: valence detection and like/dislike detection. Recognition accuracy of 77.5% and 78.0%, respectfully, was achieved. The average accuracy obtained ranges from 70.8% to 78%. The analysis of the experimental results leads to the conclusion that the proposed new type of descriptor achieves high classification accuracy that is comparable to that of other established descriptors. Therefore, the decorated energy of the time-space distribution of EEG activity can be successfully used in solving problems related to the classification of negative emotional states.

[B4.8]. T. Kostoulas, T. Winkler, **T. Ganchev**, N. Fakotakis, J. Köhler (2013). "The MoveOn Database: A Motorcycle Environment Speech and Noise Database for Command and Control Applications," *Language Resources and Evaluation Journal*, ISSN: 1574-020X, Springer, vol.47, no.2, pp.539-563. (**Thomson Reuters IF2013 = 0.518**)

The MoveOn speech and noise database was purposely designed and implemented in support of research on spoken dialogue interaction in a motorcycle environment. The distinctiveness of the MoveOn database results from the requirements of the application domain—an information support and operational command and control system for the two-wheel police force—and also from the specifics of the adverse open-air acoustic environment. In this article, we first outline the target application, motivating the database design and purpose, and then report on the implementation details. The main challenges related to the choice of equipment, the organization of recording sessions, and some difficulties that were experienced during this effort, are discussed. We offer a detailed account of the database statistics, the suggested data splits in subsets, and discuss results from automatic speech recognition experiments, which illustrate the degree of complexity of the operational environment. The detailed statistical description of the database, includes (i) annotations of speech, (ii) the emotional coloration and stress levels in speech, (iii) the typical categories of acoustic events, and (iv) the typical categories noises from the natural environment. We provide dividing the data into subsets, necessary for the efforts for creation of acoustic models and their experimental evaluation.

The analysis of results of experimental studies of precision and accuracy of automatic speech recognition in conditions, characterized by the combination of physical and cognitive stress. In such a way we illustrate the complexity of the particular operating environment in which the police motorcycle division operates.

This work was supported by the FP6 MoveOn project (IST-2005-034753), which was co-funded by the European Commission.

[B4.9]. T. Kostoulas, I. Mporas, O. Kocsis, **T. Ganchev** , N. Katsaounos, J. J. Santamaria, S. Jimenez-Murcia, F. Fernandez-Aranda, N. Fakotakis (2012). "Affective Speech Interface in Serious Games for Supporting Therapy of Mental Disorders", Expert *Systems with Applications,* vol.39, no.12, September 2012, pp.11072-11079. **(Thomson Reuters IF2012 = 1.854),** DOI = 10.1016 / j.eswa.2012.03.067.

We describe a novel design, implementation and evaluation of a speech interface, as part of a platform for the development of serious games. The speech interface consists of the speech recognition component and the emotion recognition from speech component. The speech interface relies on a platform designed and implemented to support the development of serious games, which supports cognitive-based treatment of patients with mental disorders. The implementation of the speech interface is based on the Olympus/RavenClaw framework. This framework has been extended for the needs of the specific serious games and the respective application domain, by integrating new components, such as emotion recognition from speech.

The evaluation of the speech interface utilized purposely-collected domain-specific dataset. The speech recognition experiments show that emotional speech moderately affects the performance of the speech interface. Furthermore, the emotion detectors demonstrated satisfying performance for the emotion states of interest, Anger and Boredom, and contributed towards successful modelling of the patient's emotion status.

The performance achieved for speech recognition and for the detection of the emotional states of interest was satisfactory. Recent evaluation of the serious games showed that the patients started to show new coping styles with negative emotions in normal stress life situations.

This work was supported by the PlayMancer project (FP7-ICT-215839-2007), which was co-funded by the Seventh Framework Programme of the European Commission and CIBER (initiative of Instituto Salud Carlos III). The authors wish to thank the European Commission and CIBER, as well as all members of the project consortium for their support.

[B4.10]. F. Fernandez-Aranda, S. Jimenez-Murcia, J. J. Santamaria, K. Gunnard, A. Soto, E. Kalapanidas , R. G. Bults, C. Davarakis, **T. Ganchev** , R. Granero, D. Konstantas, T. Kostoulas, T. Lam, M. Lucas, C. Masuet-Aumatell, M. H. Moussa, J. Nielsen, and E. Penelo (2012). Video games as a complementary tool in mental disorders: Playmancer a European multicenter study, *Journal of Mental Health,* 2012 Informa UK, Ltd., ISSN: 0963-8237, ISSN: 1360-0567, August 2012, vol.21, issue 4, pp. 364–374. DOI = 10.3109 / 09638237.2012.664302. (**Thomson Reuters IF2012 = 1.389**)

In article [B4.10] we present the concept and main results of the implementation of the highly regarded European research project PlayMancer, which designs, develops and evaluates the feasibility of a newly created serious video game designed to aid recovery in patients with behavioral and emotional disorders, addictions and other mental issues. The video game uses biofeedback to help patients learn relaxation skills, learn self-control strategies, and come up with new emotional regulation strategies. The article provides a brief description of the video game developed, the rationale, user requirements, innovative technology for multimodal recognition of emotional states and preliminary outcomes, regarding the utility in the treatment of certain mental disorders (gambling addiction, eating disorders). In short, research has shown that the video game created has the potential and capacity to change underlying attitudes, behaviors, and emotional processes in patients with impulsive disorders. The game provides opportunities for new modes of interaction and an objective assessment of the patient's emotional state. This is mainly due to the implemented new functionalities, using innovative methods for recognizing emotions from speech, facial expression and physiological signals.

Based on a database of speech, video, and physiological signals (including ECG, skin surface resistance, respiratory rate, etc.) specifically designed for the purpose of this study, statistical models of emotional states have been created, as and new methods for multimodal recognition of emotions and specific mental states. These innovative methods for multi-modal recognition of emotional states, integrated in the video game, in order to create biofeedback, allow logic and strategy adaptation depending on the current emotional and psychological state of the player. This creates a direct link between the player's real-world status and the on-screen character's behavior in the protected world of the video game. Thus, the use of biofeedback creates the conditions to support the acquisition of skills for relaxation, self-control and emotional regulation.

While there are previous studies, showing the usefulness of serious games in the treatment of certain behaviors, or additional agents in the treatment of diseases in several areas (schizophrenia, asthma or limb rehabilitation), the innovative methodology and technological development fills the lack of research and the possibility of dedicated serious games, to be used to support the treatment of mental disorders.

Support was given by the PlayMancer project (FP7-ICT-215839-2007), which was funded by the FP7 of the European Commission. The project also received partial support from ISCIII (CIBER06/03, FIS PI081573). CIBERobn is an ISCIII initiative.

[B4.11]. T. Kostoulas, **T. Ganchev**, A. Lazaridis, N. Fakotakis (2010). "Enhancing Emotion Recognition from Speech through Feature Selection", Proc. of the 13th International Conference on Text, Speech and Dialogue, TSD 2010, Brno, Czech Republic, September 6-10, 2010. *Lecture Notes in Computer Science,* pp.338–344. DOI = 10.1007 / 978-3-642-15760-8_43

In the present work, we aim at performance optimization of a speaker independent emotion recognition system through speech feature selection process. Specifically, relying on the speech feature set defined in the Interspeech 2009 Emotion Challenge, we studied the relative importance of the individual speech parameters, and based on their ranking, a subset of speech parameters that offered advantageous performance was selected. The affect-emotion recognizer utilized here relies on a GMM-UBM-based classifier. In all experiments, we followed the experimental setup defined by the Interspeech 2009 Emotion Challenge, utilizing the FAU Aibo Emotion Corpus of spontaneous, emotionally colored speech. The experimental results indicate that the correct choice of the speech parameters can lead to better performance than the baseline one.

The main contribution of this research is the systematic evaluation of the relative importance of the individual speech descriptors and their relevance to the task of recognizing emotional states. For this purpose, the basic item set of descriptors 384 is subjected to systematic evaluation method for search BestFirst at ten-fold evaluation of the quality of individual descriptors. For the purposes of this study, 10 different subsets of the development dataset were used. After ten times cross-validation the assessment of relevance selected 56 speech features, including only those, which have been selected at least 5 times in the ten tests. These 56 descriptors are sorted on the basis of their importance, and then subsets are compiled, including only the first, first two, first three, etc. A total of 56 experiments were conducted to evaluate the accuracy of recognition of the five emotion categories, for each subset. Based on the analysis of the experimental results, a set was selected that included only the first 49 descriptors out of 56; the addition of the following does not lead to higher accuracy. We followed the experimental protocol defined by Interspeech 2009 Emotion Challenge. A database of spontaneous emotionally-colored FAU Aibo Emotion Corpus speech was used. It contains speech from 51 children collected while playing with an Aibo robot, commanding the robot to perform predefined action sequences designed to provoke emotional reactions.

Analysis of the experimental results shows that the proposed method for selecting a set of descriptors contributes to improving the accuracy of emotional states detection with respect to the baseline subset. At the same time, the required computing power and data processing time are reduced.

This work was supported by the PlayMancer project (FP7-ICT-215839-2007), which is funded by the Seventh Framework Programme of the European Commission.

[B4.12]. O. Kocsis, **T. Ganchev** , I. Mporas, G. Papadopoulos, N. Fakotakis: "*Multi-modal System Architecture for Serious Gaming*", Proc. published in the IFIP International Federation for Information Processing, Volume 296; in " *Artificial Intelligence Applications and Innovations III*"; Eds. Iliadis, L., Vlahavas, I., Bramer, M.; (Boston: Springer), pp.441-447.

Human-computer interaction (HCI), especially in the games domain, targets to mimic as much as possible the natural human-to-human interaction, which is multimodal, involving speech, vision, haptic, etc. Furthermore, the domain of serious games, aiming to value-added games, makes use of additional inputs, such as biosensors, motion tracking equipment, etc. In this context, game development has become complex, expensive and burdened with a long development cycle. This creates barriers to independent game developers and inhibits the introduction of innovative games, or new game genres. In this paper, the PlayMancer platform is introduced, a work in progress aiming to overcome such barriers by augmenting existing 3D game engines with innovative modes of interaction. Playmancer integrates open source existing systems, such as a game engine and a spoken dialog management system, extended by newly implemented components, supporting innovative interaction modalities, such as emotion recognition from audio data, motion tracking, etc. and advanced configuration tools.

This work was supported by the PlayMancer project (FP7 215839), which is partially funded by the European Commission.

[B4.13]. T. Kostoulas, **T. Ganchev**, I. Mporas, N. Fakotakis (2007). "Detection of Negative Emotional States in Real-World Scenario", *Proc. of the 19th IEEE International Conference on Tools with Artificial Intelligence*, ICTAI-2007. October 29-31, 2007, Patras, Greece, pp.502-509. DOI = 10.1109/ictai.2007.106

In the present work, we evaluate a detector of negative emotional states (DNES) that serves the purpose of enhancing a spoken dialogue system, which operates in smart-home environment. The DNES component is based on Gaussian mixture models (GMMs) and a set of commonly used speech features. In comprehensive performance evaluation, we utilized a well-known acted speech database and real-world speech recordings. The real-world speech was collected during interaction of naïve users with our smart-home spoken dialogue system. The experimental results show that the accuracy of recognizing negative emotions on the real world data is lower than the one reported when testing on the acted speech database, though much promising, considering that, often, humans are unable to distinguish the emotion of other humans judging only from speech.

This work was partially supported by the LOGOS project, which is funded by the General Secretariat for Research and Technology of the Greek Ministry of Development.

7.2 Abstracts of the publications grouped by criterion ”G.7”

[G7.1]. N. Dukov, **T. Ganchev,** and D. Kovachev (2017). "FPGA implementation of the Locally Recurrent Probabilistic Neural Network". Proc. of the IITI-2017, September 14-16, 2017, Varna, Bulgaria, in “Advances *in Intelligent Systems and Computing* ", ISSN: 2194-5357, vol. 680, pp.419-428.

The Locally Recurrent Probabilistic Neural Network (LRPNN) consists of an input layer, three hidden layers and an output layer. The first two hidden layers are derived from the original PNN, while the third layer referred as recurrent layer is capable to model correlations within temporal sequences of observations. In the present study, we investigate the feasibility of FPGA-based implementation of the locally recurrent layer of LRPNN. An important consideration due to the specifics of this architecture is the use of modules with very high precision in the hardware design. Although expensive in terms of available resources in the FPGA chip, this is necessary, in order to compensate for the added error of quantization due to the multiple feedbacks from neurons in the neural network. The weights for the recurrent layer of the LRPNN are automatically computed from the available training data and translated into the hardware design.

The experimental evaluation was carried out on the DEAP database, where two classes of emotional states were considered. The design makes use of a computed short-term energy from a 32-channel electroencephalographic (EEG) signal as an input. Results of an extensive experimental validation show that there is approximately one percent difference between the accuracy achieved with CPU-based software and FPGA-based hardware implementation of the LRPNN.

The research and results reported in this study were developed with the financial support of TU-Varna, Project НП7/2017, named “Development of design capabilities and hardware implementation of monolithic integrated circuits and electronic systems with programmable analog matrices”.

[G7.2]. V. Markova, K. Kalinkov, P. Stanev, **T. Ganchev** (2017). "Automated Stress Level Monitoring in Mobile Setup". *Proc. of the IITI-2017,* September 14-16, 2017, Varna, Bulgaria, in "Advances in Intelligent Systems and Computing", ISSN: 2194-5357, vol. 680, pp.323-331.

We present the design of a mobile system for real-time stress-level assessment. The system combines wearable sensors, wireless data acquisition, and Cloud computing in order to collect and analyze physiological signals, such as, Galvanic Skin Response (GSR) and skin temperature. We report on the implementation of a specific use case, which incorporates functionality for real-time data logging and analysis.

Experimental results demonstrate excellent recognition accuracy of affective arousal and decent accuracy for binary detection of valence. In addition, we also evaluate the feasibility for detection of high arousal/negative valence (HANV) events, which in specific setups can be connected to stress.

The authors acknowledge with sincere thanks the support received through the research project NP5/2017 entitled "Study of Methods and Apparatus for the Acquisition of Biomedical Data in Mobile Setup", financed by the National Science Fund of Bulgaria and Technical University of Varna.

[G7.3]. S. Ntalampiras, D. Arsic, M. Hofmann, M. Andersson, **T. Ganchev** (2014). PROMETHEUS: heterogeneous sensor database in support of research on human behavioral patterns in unrestricted environments, *Signal, Image and Video Processing,* ISSN: 1863-1703. October 2014, vol.8, no.7, pp.1211-1231, DOI = 10.1007/s11760-012-0346-9, (**Thompson Reuters IF2014 = 1.430**)

Abstract The multi-modal multi-sensor PROMETHEUS database was created in support of research and development activities [PROMETHEUS (FP7-ICT-214901): http://www.prometheus-FP7.eu] aiming at the creation of a framework for monitoring and interpretation of human behaviors in unrestricted indoor and outdoor environments. The distinctiveness of the PROMETHEUS database comes from the unique sensor sets, used in the various recording scenarios, but also from the database design, which covers a range of real-world applications, correlated to smart-home automation and indoors/outdoors surveillance of public areas. Numerous single-person and multi-person scenarios, but also scenarios with interactions between groups of people, motivated by these applications were implemented with the help of skilled actors and supernumerary personnel. In these scenarios, the actors and personnel were instructed to implement a range of typical and atypical behaviors, and simulations of emergency and crisis situations. In summary, the database contains more than 4 h of synchronized recordings from heterogeneous sensors (an infrared motion detection sensor, thermal imaging cameras, overview/surveillance video cameras, close-view video cameras, a 3D camera, a stereoscopic camera, a general-purpose camcoder, microphone arrays, and motion capture equipment) collected in common setups, simulating smart-home environment, airport, and ATM security environment. Selected scenes of the database were annotated for the needs of human detection and tracking. The entire audio part of the database was annotated for the needs of sound event detection, sound source enumeration, emotion recognition, etc.

The work reported in this paper was done within contract FP7-214901 “Prediction and Interpretation of human behavior based on probabilistic models and heterogeneous sensors”, PROMETHEUS project, co-funded by the FP7 of the European Commission.

[G7.4]. S.M. Potirakis, N.-A. Tatlas, N. Zafeiropoulos, T. Ganchev, M. Rangoussi (2013). "Assessment of Military Intercom Heads for Maximum Voice Reproduction Level in High Noise Conditions," *Applied Acoustics*, ISSN 0003-682X, vol.74, no.6, June 2013, pp.870-881, DOI = 10.1016 / j.apacoust. 2012/12/19 (**Thomson Reuters IF2013 = 1.068**)

Intercom headsets are mandatory communication apparatus in high noise environments (HNE). The headset selection in HNE, such as combat vehicles, is crucial for achieving the objectives of communication, as it serves the needs for both noise reduction and voice reproduction. Although military-grade intercom headsets are typically used under extreme environmental conditions, a standard performance evaluation method exists only for the earphone elements. In the present work, we propose an integrated method for the assessment of the electroacoustic performance of HNE headsets in conditions of maximum reproduction level and high environmental noise, focusing on the voice communication quality. Objective methods, such as Automatic Speech Recognition (ASR), Perceptual Evaluation of Speech Quality (PESQ) and Speech Transmission Index (STI) are comparatively evaluated and their results are compared to subjective scores using Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) in order to reveal the best-fit metrics.

The authors wish to thank Intracom Defense Electronics S.A. for the kind permission to use the facilities of their Analog Electronics and Electroacoustics Laboratory for the conduction of the presented measurements, as well as for the organization of the subjective evaluation tests inside the tracked vehicle.

[G7.5]. A. Lazaridis, **T. Ganchev,** I. Mporas, E. Dermatas, N. Fakotakis, (2012). "Two-stage phone duration modeling with feature construction and feature vector extension for speech synthesis needs", *Computer Speech & Language* , ISSN 0885-2308, vol.26, no.4, August 2012, pp.274-292, DOI = 10.1016 / j.csl.2012.01.009. **(Thomson Reuters IF2014 = 1.463)**

We propose a two-stage phone duration modelling scheme, which can be applied for the improvement of prosody modelling in speech synthesis systems. This scheme builds on a number of independent feature constructors (FCs) employed in the first stage, and a phone duration model (PDM) which operates on an extended feature vector in the second stage. The feature vector, which acts as input to the first stage, consists of numerical and non-numerical linguistic features extracted from text. The extended feature vector is obtained by appending the phone duration predictions estimated by the FCs to the initial feature vector.

Experiments on the American-English KED TIMIT and on the Modern Greek WCL-1 databases validated the advantage of the proposed two-stage scheme, improving prediction accuracy over the best individual predictor, and over a two-stage scheme, which just fuses the first-stage outputs. Specifically, when compared to the best individual predictor, a relative reduction in the mean absolute error and the root mean square error of 3.9% and 3.9% on the KED TIMIT, and of 4.8% and 4.6% on the WCL-1 database, respectively, is observed.

[G7.6]. A. Lazaridis, **T. Ganchev,** T. Kostoulas, I. Mporas, N. Fakotakis (2010). "Phone Duration Modeling: An Overview of Techniques and Performance Optimization via Feature Selection in the Context of Emotional Speech," *International Journal of Speech Technology,* ISSN 1381-2416, vol.13, no.3, 2010. Springer, pp. 175-188.

Accurate modeling of prosody is prerequisite for the production of synthetic speech of high quality. Phone duration, as one of the key prosodic parameters, plays an important role for the generation of emotional synthetic speech with natural sounding. In the present work, we offer an overview of various phone duration modeling techniques, and consequently evaluate ten models, based on decision trees, linear regression, lazy-learning algorithms and meta-learning algorithms, which over the past decades have been successfully used in various modeling tasks. Furthermore, we study the opportunity for performance optimization by applying two feature selection techniques, the RReliefF and the Correlation-based Feature Selection, on a large set of numerical and nominal linguistic features extracted from text, such as: phonetic, phonologic and morpho-syntactic ones, which have been reported successful on the phone and syllable duration modeling task. We investigate the practical usefulness of these phone duration modeling techniques on a Modern Greek emotional speech database, which consists of five categories of emotional speech: anger, fear, joy, neutral, sadness.

The experimental results demonstrated that feature selection significantly improves the accuracy of phone duration prediction regardless of the type of machine learning algorithm used for phone duration modeling. Specifically, in four out of the five categories of emotional speech, feature selection contributed to the improvement of the phone duration modeling, when compared to the case without feature selection. The M5p trees based phone duration model was observed to achieve the best phone duration prediction accuracy in terms of RMSE and MAE.

[G7.7]. I. Mporas, **T. Ganchev,** O. Kocsis, N. Fakotakis (2009). "Performance Evaluation of a Speech Interface for the Motorcycle Environment", 5th IFIP Conference on Artificial Intelligence Applications & Innovations (AIAI 2009), April 23-25, 2009, Proc. published in the IFIP International Federation for Information Processing, vol.296; ISBN: 978-1-441-90220-7, "*Artificial Intelligence Applications and Innovations*" III; Eds. Iliadis, L., Vlahavas, I., Bramer, M.; (Boston: Springer), pp.259-266.

In the present work, we investigate the performance of a number of traditional and recent speech enhancement algorithms in the adverse non-stationary conditions, which are distinctive for motorcycle on the move. The performance of these algorithms is ranked in terms of the improvement they contribute to the speech recognition rate, when compared to the baseline result, i.e. without speech enhancement.

The experimentations on the MoveOn motorcycle speech and noise database suggested that there is no equivalence between the ranking of algorithms based on the human perception of speech quality and the speech recognition performance. The Multi-band spectral subtraction method was observed to lead to the highest speech recognition performance.

This work was supported by the MoveOn project (IST-2005-034753), which is partially funded by the European Commission.

[G7.8]. A. Conconi, **T. Ganchev,** O. Kocsis, G. Papadopoulos, F. Fernández-Aranda, S. Jiménez-Murcia (2008). PlayMancer: A Serious Gaming 3D Environment, *Proc of the 4th International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution, AXMEDIS*- 2008, Florence, Italy, Nov. 17-19, 2008.

Serious games are computer games used as educational technology or as a vehicle for presenting or promoting a point of view. Serious games are intended to provide an engaging, self-reinforcing context in which to motivate and educate the players towards non-game events or processes, including business operations, training, marketing and advertisement. The potential of games for entertainment and learning has been demonstrated thoroughly from research and clearly in the market place. Unfortunately, the investments committed to entertainment dwarfs that which is committed for more serious purposes. Furthermore, game development has become more complex, expensive, and burdened with a long development cycle.

In this paper, we introduce PlayMancer, a work in progress that aims to overcome such barriers by augmenting existing 3D gaming engines with new possibilities and thusly creating a novel development framework. In section I of this paper we briefly survey the serious games market. In section II, we introduce the PlayMancer project and its objectives, whereas in section III we present the platform architecture. Section VI describes an application scenario where a PlayMancer-based game is being used as additional therapeutic tool to treat chronic mental disorders, such as eating disorders and behavioral addiction.

This work was developed within the PlayMancer project (www.playmancer.eu), which is partially funded by the European Commission under the Seventh Framework Programme (FP7-ICT-215839- 2007). The authors wish to thank the Commission as well as all members of the project consortium for their support.

[G7.9]. I. Mporas, **T. Ganchev,** N. Fakotakis (2008). "Phonetic Segmentation Using Multiple Speech Features", *International Journal of Speech Technology* , ISSN 1381-2416, vol.11, no.2, June 2008, pp.73-85.

In this paper, we propose a method for improving the performance of the segmentation of speech waveforms to phonetic units. The proposed method is based on the well-known Viterbi time-alignment algorithm and utilizes the phonetic boundary predictions from multiple speech parameterization techniques. Specifically, we utilize the most appropriate, with respect to boundary type, phone transition position prediction as initial point to start Viterbi time-alignment for the prediction of the successor phonetic boundary.

The proposed method was evaluated on the TIMIT database, with the exploitation of several, well known in the area of speech processing, Fourier-based and wavelet-based speech parameterization algorithms.

The experimental results for the tolerance of 20 milliseconds indicated an improvement of the absolute segmentation accuracy of approximately 0.70%, when compared to the baseline speech segmentation scheme.

[G7.10]. I. Mporas, **T. Ganchev,** O. Kocsis, N. Fakotakis (2010). "Speech Enhancement for Robust Speech Recognition in the Motorcycle Environment", *International Journal of Artificial Intelligence Tools* (IJAIT), ISSN: 0218-2130, *Special issue on "Artificial Intelligence Techniques for Pervasive Computing",* vol.19, no.2, 2010. pp. 159-173. **(Thomson Reuters IF2010 = 0.330)**

In the present work, we investigate the performance of a number of traditional and recent speech enhancement algorithms in the adverse non-stationary conditions, which are distinctive for motorcycles on the move. The performance of these algorithms is ranked in terms of the improvement they contribute to the speech recognition accuracy, when compared to the baseline performance, i.e. without speech enhancement.

The experiments on the MoveOn motorcycle speech and noise database indicated that there is no equivalence between the ranking of algorithms based on the human perception of speech quality and the speech recognition performance. The Multi-band spectral subtraction method was observed to lead to the highest speech recognition performance.

This work was partially supported by the MoveOn project (IST-2005-034753) (http://www.m0ve0n.net/), which is co-funded by the FP6 of the European Commission.

[G7.11]. A. Lazaridis, I. Mporas, **T. Ganchev,** N. Fakotakis (2011). "Support Vector Regression Fusion Scheme in Phone Duration Modeling", *Proc. of the 2011 IEEE International Conference on Acoustics, Speech, and Signal Processing,* ICASSP-2011, Prague, Check Republic, pp. 4732-4735.

A fusion scheme of phone duration models (PDMs) is presented in this work. Specifically, a support vector regression (SVR)-fusion model is fed with the predictions of a group of independent PDMs operating in parallel. The American-English KED TIMIT and the Greek WCL-1 databases are used for evaluating the PDMs and the fusion scheme.

The fusion scheme contributes to the accuracy improvement over the best individual model, achieving a relative reduction of the mean absolute error (MAE) and the root mean square error (RMSE), by 1.9% and 2.0% on KED TIMIT, and 2.6% and 1.8% respectively on WCL-1.

Moreover, for evaluating the impact the accuracy improvement will have on synthetic speech, perceptual evaluation test was performed. This test showed that the accuracy improvement achieved by the SVR-fusion would contribute to the improvement of the naturalness of synthetic speech.

[G7.12]. A. Lazaridis, T. Kostoulas, **T. Ganchev,** I. Mporas, N. Fakotakis (2010). "Vergina: A Modern Greek Speech Database for Speech Synthesis, *" Proc. of the Seventh Conference on International Language Resources and Evaluation* , LREC-2010 , May 19-21, Valletta, Malta, Eds. N. Calzolari et al., European Language Resources Association (ELRA), ISBN: 2-9517408-6-7, pp.117-121.

The present paper outlines the Vergina speech database, which was developed in support of research and development of corpus-based unit selection and statistical parametric speech synthesis systems for Modern Greek language. In the following, we describe the design, development and implementation of the recording campaign, as well as the annotation of the database. Specifically, a text corpus of approximately 5 million words, collected from newspaper articles, periodicals, and paragraphs of literature, was processed in order to select the utterances-sentences needed for producing the speech database and to achieve a reasonable phonetic coverage. The broad coverage and contents of the selected utterances-sentences of the database – text corpus collected from different domains and writing styles – makes this database appropriate for various application domains. The database, recorded in audio studio, consists of approximately 3,000 phonetically balanced Modern Greek utterances corresponding to approximately four hours of speech. Annotation of the Vergina speech database was performed using task-specific tools, which are based on a hidden Markov model (HMM) segmentation method, and then manual inspection and corrections were performed.

[G7.13]. I. Mporas, A. Lazaridis, **T. Ganchev,** N. Fakotakis (2009). "Using Hybrid HMM-Based Speech Segmentation to Improve Synthetic Speech Quality," *Proc of the 13th Panhellenic Conference in Informatics,* PCI-2009, Corfu, Greece, September 10-12, 2009, pp.118-122.

The automatic phonetic time-alignment of speech databases is essential for the development cycle of a Text-to- Speech (TTS) system. Furthermore, the quality of the synthesized speech signals is strongly related to the precision of the produced alignment. In the present work, we study the performance of a new HMM-based speech segmentation method. The method is based on hybrid embedded and isolated-unit trained models, and has proved to improve the phonetic segmentation accuracy in the multiple speaker task. Here it is employed on the single speaker segmentation task, utilizing a Greek-speech database.

The evaluation of the method showed significant improvement in terms of phonetic segmentation accuracy as well as in the perceptual quality of synthetic speech, when compared to the baseline system.

This work was supported by the MoveOn project (IST-2005-034753), which is co-funded by the FP6 of the European Commission.

[G7.14]. I. Mporas, **T. Ganchev,** T. Kostoulas, K. Kermanidis, N. Fakotakis (2009). "Automatic Speech Recognition System for Home Appliances Control", *Proc of the 13th Panhellenic Conference in Informatics,* PCI-2009, Corfu, Greece, September 10-12, 2009, pp.114-117.

In the present work, we study the performance of a speech recognizer for the modern Greek language, in a smart-home environment. This recognizer operates in spoken interaction scenarios, where the users are able to control various home appliances. In contrast to command and control systems, in our application the users speak spontaneously, beyond the use of a standardized set of isolated commands.

The speech recognition system was developed using open source software, the HMM ToolKit (HTK), which runs on any home computer. Using HTK and database SpeechDat (II) -FDB-5000-Greek we created an acoustic model, which uses HMM modeling with 5 states in the topology of Bakis. The grammar contains 73 words -- commands and their synonyms, the most commonly used words, for voice control of TV, telephone, DVD player, washing machine, etc. and includes a common model (*garbage model*), which is representative of the spontaneous speech, which is not of interest for the management of home appliances.

The performance of the automatic speech recognition system has been tested for 10 users with different noise characteristics of the environment, for two different types of microphones -- a regular PC microphone and a high-quality AKG C444 wireless microphone, which is equipped with an automatic noise canceling method.

During the experimental validation of the system under different signal-to-noise ratios (SNR 6dB, 12 dB, 30 dB) and high-quality speech, using the AKG C444 microphone, the command recognition precision is in the range of 94-96% and 98 %, respectively. Despite the different word recognition error rates, both microphones reported 100% task completion in all scenarios .

The use of open source speech recognition in combination with inexpensive microphone for close talking, enables cheaper solution for home automation system. The user receives feedback simply by monitoring the operational status of the device.

[G7.15]. I. Mporas, O. Kocsis, **T. Ganchev,** N. Fakotakis (2009). "A Collaborative Speech Enhancement Approach to Speech Recognition In a Motorcycle Environment," *Proc. of the 16th International Conference on Digital Signal Processing,* DSP 2009, Santorini, Greece. July 5-7, 2009. Article number 167.

Aiming at the optimization of the speech recognition performance, we investigate various configurations for a speech front-end, which is part of a multimodal dialogue interaction interface of a wearable solution for information support of the motorcycle police force on the move. Initially, the practical value of various speech enhancement techniques is assessed, and subsequently a collaborative scheme employing independent speech enhancement channels, which operate in parallel on a common input, is proposed.

It was experimentally found that the Adaboost.M1 algorithm is the most advantageous among a number of fusion methods. The improvement of speech recognition accuracy due to the collaborative speech enhancement scheme is expressed as gain of 8 % in terms of word recognition rate, when compared to the performance of the best speech enhancement channel, alone.

This work was supported by the MoveOn project (IST2005-034753), which is partially funded by the European Commission.

[G7.16]. R. Saeidi, HS Mohammadi, **T. Ganchev,** RD Rodman (2008). "Effects of Feature Domain Normalizations on Text-Independent Speaker Verification Using Sorted Adapted Gaussian Mixture Models," *Proc. of the 13th International CSI Computer Conference,* CSICC-2008, Kish Island, Iran, March 9-11, 2008.

In this paper, we evaluate sorted Gaussian Mixture Model (GMM) system performance for Text Independent Speaker Verification under the feature-domain normalization conditions. Sorted GMM is a speed-up algorithm proposed for GMM based systems. Cepstral Mean Subtraction (CMS) and Dynamic Range Normalization (DRN) are the normalization schemes studied for sorted GMM system purposes. Effectiveness of these normalizations has been proved in speaker recognition systems while their effectiveness on the speed-up of GMM based speaker verification is showed in this study.

The baseline system is a universal background model–Gaussian mixture model (UBM-GMM) system and evaluations were performed on the NIST 2002 speaker recognition evaluation database with NIST SRE rules.

It is shown that CMS and DRN normalizations enhance both the baseline system and sorted GMM system performances. In other words, the performance loss due to reducing the computational load is mitigated by applying CMS and DRN.

[G7.17]. R. Saeidi, HRS Mohammadi, **T. Ganchev,** RD Rodman (2008). "Hierarchical mixture clustering and its application to GMM-based text independent speaker identification," *Proc. of the International Symposium on Telecommunications,* IST2008, Tehran, Iran, Aug. 27-28, 2008, pp.770-773.

In this paper, we propose a hierarchical mixture clustering method and investigate its application for complexity reduction of a GMM based speaker identification system. We show that by using GMM-HMC, we can cluster speakers more accurately than that of a sorted GMM with the same acceleration rate.

The system was tested on a universal background model–Gaussian mixture model with KL-divergence as the distance measure. While the proposed system's performance is inferior to the baseline system, its comparatively smaller computational load provides the potential to develop systems with higher performance.

The analysis of experimental results of the comparative evaluation shows that by using GMM-HMC it is possible to achieve a more accurate grouping of the speakers than with a method using a sorted GMM with the same degree of acceleration. The system has been tested for the basic GMM-HMC and three modified variants. In all cases, the methods for reducing the number of calculations reduce the accuracy of speaker identification by 3% - 5% relative to the base system by reducing the number of calculations by about 3-4 times. The analysis of the results showed a loss of accuracy of about 3.3% with a decrease in the number of calculations 3.2 times.

Based on the experimental results, it was concluded that further development of the method of hierarchical clustering (GMM-HMC), has the potential for the creation of systems for speaker identification with higher accuracy and lower computational demands.

[G7.18]. I. Mporas, **T. Ganchev,** N. Fakotakis (2008). "A Hybrid Architecture for Automatic Segmentation of Speech Waveforms," *Proc. of the 2008 IEEE International Conference on Acoustics, Speech, and Signal Processing* , ICASSP-2008, Las Vegas, USA, March 30 - April 4, 2008, pp.4457-4460.

In the present work, we propose a hybrid architecture for automatic alignment of speech waveforms and their corresponding phone sequence. The proposed architecture does not exploit any phone boundary information. Our approach combines the efficiency of embedded training techniques and the high performance of isolated-unit training. Evaluating on the established for the task of phone segmentation TIMIT database, we achieved an accuracy of 83.56%, which corresponds to improving the baseline system's accuracy by 6.09 %.

The proposed new approach is illustrated by a new automated method for segmentation of the speech, which combines the effectiveness of integrated methods for training of phonemic patterns generated by Hidden Markov Model (HMM) and high accuracy isolated training of HMM. As the proposed new method does not use manually created phonetic transcriptions, all HMM models are initialized uniformly, with the same parameter values. The parameters of all HMM models are then recalculated and adjusted simultaneously using the Baum-Welch algorithm. This is repeated a number of times until the convergence of the model parameters is reached.

Using a standard experimental protocol, based on the TIMIT Speech Database which contains speech recordings of 640 people, segmentation accuracy was evaluated. In the environment of many different speakers, the precision of the proposed new method is compared with that of a widely used basic method. Different parameter configurations of the proposed new method are investigated, including HMM modeling with 5 and 6 states for each phoneme, which are the most widely used models in segmentation or speech recognition systems. A comparison was made with the traditional baseline method for the same number of states and using embedded and isolated model training. The analysis of the experimental results shows that after a small number of iterations (usually between 15 and 20), the precision of segmentation of the speech signal increases by more than 6%.

This work was partially supported by the LOGOS project (ΕΗΓ-102), which is funded by the General Secretariat for Research and Technology of the Greek Ministry of Development.

[G7.19]. S. Ntalampiras, I. Potamitis, **T. Ganchev,** and N. Fakotakis (2008). "Audio Database in Support of Potential Threat and Crisis Situation Management," *Proc. of the Sixth International Conference on Language Resources and Evaluation,* LR EC- 2008, Morocco, May 28-30, 2008. pp. 1288 - 1291.

This paper describes a corpus consisting of audio data for automatic space monitoring based solely on the perceived acoustic information. The particular database is created as part of a project aiming at the detection of abnormal events, which lead to life-threatening situations or property damage. The audio corpus is composed of vocal reactions and environmental sounds that are usually encountered in atypical situations. The audio data is composed of three parts:

- Phase I - professional sound effects collections,
- Phase II recordings obtained from action and drama movies and
- Phase III - vocal reactions related to real-world emergency events as retrieved from television, radio broadcast news, documentaries etc.

The annotation methodology is given in details along with preliminary classification results and statistical analysis of the dataset regarding Phase I. The main objective of such a dataset is to provide training data for automatic recognition machines that detect hazardous situations and to provide security enhancement in public environments, which otherwise require human supervision.

This work was supported by the EC FP 7th grant Prometheus 214901 “Prediction and Interpretation of human behavior based on probabilistic models and heterogeneous sensors”.

[G7.20]. M. Siafarikas, **T. Ganchev,** N. Fakotakis, G. Kokkinakis (2005). "Overlapping Wavelet Packet Features for Speaker Verification," *Proc. of the 9th European Conference on Speech Communication and Technology,* InterSpeech-2005, September 4-8, 2005, Lisbon, Portugal, pp. 3121-3124.

A generalization of the Discrete Wavelet Packet Transform (DWPT), referred to as Overlapping Discrete Wavelet Packet Transform (ODWPT), is proposed. In contrast to the traditional DWPT, the ODWPT assumes overlapping among the frequency sub-bands at various levels of the transform. Based on this overlapping strategy, a new set of speech features that is specially designed for speaker recognition is derived.

The practical significance of our approach has been evaluated in comparative experiments performed on the 2001 NIST Speaker Recognition Evaluation database. To study the suitability of the new set of descriptors (WP-2011), a standard experimental protocol, defined by the organizers of the 2001 NIST SRE competition, was followed. For this, we used the dataset recorded in the mobile phone networks of the United States. Following standard experimental protocol for one-speaker detection, we performed a comparative analysis between the proposed descriptors, WP-2011 and other five types of descriptors. Among these are two recently created (WP2-Sarikaya and WP3-Farooq-Data), using the transformation with wavelet packets, two traditional (MFCC FB-20 (HTK) and MFCC FB-32) using time-frequency analysis with short-term Discrete Fourier Transform and a features denoted as WP-0000, in which wavelet descriptors are calculated using orthogonal set of basic functions.

The analysis of the experimental results with the proposed new type of speech descriptors (WP-2011), which are optimized for speaker verification purposes shows a relative reduction in recognition error of 27%, 22% and 15% relative to the WP3-Farooq-Data, MFCC descriptors FB-32, and WP2-Sarikaya. The use of non-orthogonal wavelet packet conversion contributes to a relative 3% reduction in speaker verification error over the equivalent orthogonal wavelet transform.

[G7.21]. M. Maragoudakis, **T. Ganchev,** N. Fakotakis (2004). "Bayesian Reinforcement for a Probabilistic Neural Net Part-Of-Speech Tagger", *Lecture Notes in Computer Science*, Springer-Verlag, Heidelberg, ISSN: 0302-9743, vol. LNAI 3206/2004, Sept ember 2004, pp.137-145.

The present paper introduces a novel stochastic model for Part-Of-Speech tagging of natural language texts. While previous statistical approaches, such as Hidden Markov Models, are based on theoretical assumptions that are not always met in natural language, we propose a methodology, which incorporates fundamental elements of two distinct machine-learning disciplines.

We make use of Bayesian knowledge representation to provide a robust classifier, namely a Probabilistic Neural Network one, with additional context information in order to better infer on the correct Part-Of-Speech label. As for training material, we make use of minimal linguistic information, i.e. only a small lexicon that contains the words that belong to non-declinable POS categories and closed-class words. Such minimal information is augmented by statistical parameters generated by Bayesian networks learning and the outcome is fed into the Probabilistic Neural Network classifier for the task of Part-Of-Speech tagging. Experimental results portray satisfactory performance, in terms of 3.5%–4% error rate.

[G7.22]. M. Siafarikas, **T. Ganchev,** N. Fakotakis (2004). "Objective Wavelet Packet Features for Speaker Verification", *Proc. 8th International Conference on Spoken Language Processing* , InterSpeech-2004 - ICSLP, Oct 4- 8, Jeju, Korea, 2004, pp. 2365-2368.

Studying ways for achieving a better demarcation of human voices for the task of speaker verification and taking advantage of the flexibility provided by wavelet packet analysis, we investigate in an objective way the relative importance of constituent disjoint frequency subbands of speech signals. Based on experimental results measuring the actual contribution of these subbands in relation to the corresponding frequency resolution, we propose a novel wavelet packet-based speech feature set that is effectively designed for speaker verification.

The practical significance of our approach has been evaluated in comparative experiments performed on 2001 NIST Speaker Recognition Evaluation database. The proposed wavelet packet feature set has proven to outperform the widely used Mel-frequency scaled cepstral coefficients (MFCCs), as well as other wavelet packet based features that have been successfully used for speaker recognition.

This work was supported by the “Infotainment management with Speech Interaction via Remote microphones and telephone interfaces” - INSPIRE project (IST-2001-32746).

7.3 Abstracts of the publications grouped by criterion ”G.8”

[G8.1]. N. Dukov, **T. Ganchev** (2018). "Novel ReLU-based neuron models for the LRPNN", *Computer Science and Technology*, vol. XVI, no.1/2018, Technical University of Varna, pp.135-142.

The locally recurrent probabilistic neural network (LRPNN) traditionally uses a sigmoidal activation function in the recurrent layer neurons. A disadvantage of this function is the computational complexity and challenging implementation in FPGA designs. In the current study, we investigate alternatives based on the ReLU activation function and its modifications.

We propose a new member of the ReLU family, referred to as SatReLU, which combines the advantages of Leaky ReLU and Clipped ReLU. The experimental validation on the negative emotion recognition task, carried out on the DEAP database, confirms the advantages of the proposed neuron models.

A comparative experimental study of the operability of the modified locally recurrent probabilistic neural network (LRPNN) was performed for different models of neurons in the locally recurrent layer. Models of neurons were studied with: (i) sigmoidal, (ii) ReLU, (iii) SatReLU, (iv) Leaky ReLU, (v) Clipped ReLU activation function .

Analysis of the experimental results show that using the proposed activation function for the neurons of the locally-recurrent layer brings advantages in terms of improved classification accuracy, reduced complexity of calculations, and an improved learning rate of the recurrent layer. Improved classification accuracy was obtained for Leaky ReLU with const = 0.07 and the proposed SatReLU with const = 80.

The SatReLU activation function performs better in terms of predictability than the other features tested, which, in view of the small difference (≈1%) with Leaky ReLU in accuracy, makes it a good alternative.

[G8.2]. F. Feradov, **T. Ganchev** (2015). "Ranking of EEG Time-Domain Features on the Negative Emotions Recognition Task", *Annual Journal of Electronics*, ISSN 1314-0078, 2015, vol. 9, pp. 26-29.

The accuracy of automated emotion recognition depends on the quality of EEG signal descriptors. In the present contribution, we report on an experimental evaluation of ten time-domain EEG signal descriptors with respect to their applicability to the task of negative emotions recognition. The ranking of these descriptors based on their estimated practical worth shows that the mean of the absolute values of the first difference of the normalized signal contributes for the highest recognition accuracy.

The authors acknowledge with thanks the support received through the research projects NP8 "Development of advanced methods for automated analysis of EEG signals for detection of negative emotional states and neurological disorders" and SNP2 "Technological support for improving the quality of life of people with the Alzheimer disease" – both financed by the Technical University of Varna, Bulgaria.

[G8.3]. Dukov N., **T. Ganchev,** D. Kovachev, M. Vrahatis (2015). "Population Size Trade-Offs in DE and PSO-Based Methods for PNN Training," *Proc. of the International Scientific Conference,* UNITECH-2015, 21-22 November, 2015, Gabrovo.

We report on a comparative evaluation of three evolutionary methods for training the probabilistic neural network (PNN). The specific focus here is on an investigation of the acceptable trade-offs, in terms of accuracy vs. computational and memory demands, depending on the population size. An empirical evaluation is carried out on the well-known Parkinson Speech Dataset with Multiple Types of Sound Recordings following a common experimental protocol. The numerical results identify the Unified PSO-based training as the most appropriate due to its superior classification accuracy and lower computational demands.

The authors acknowledge with thanks the support received through the research projects PD5 "Study of biologically substantiated architectures of artificial neural networks for the identification of heart diseases and neurological dis-orders", SNP2 "Technological support for improving quality of life of people with the Alzheimer disease", and the NP4 "Capacity building for object-oriented FPGA design in support of knowledge-based economy" financed by the Technical University of Varna, Bulgaria.

[G8.4]. Paunov P., **T. Ganchev** (2014). "Low-complexity method for height estimation from speech, *"Acoustics"*, ISSN 1312-4897, Sofia, Bulgaria, issue 16, December, 2014, pp. 121-124.

Aiming at the integration of automated height estimation in speech-based mobile applications, we compare the performance of four low-complexity implementations in terms of calculation time and mean squared error (MSE). These implementations are based on General Regression Neural Network (GRNN), feed-forward Multi-layer Perceptron Neural Network (MLP NN)-based regression, and Support Vector Regression (SVR) with Radial Basis Function (RBF) kernel and with polynomial kernel. The experimental evaluation was carried out in a common experimental setup based on the TIMIT database. We report that the MLP-based implementation provides the best trade-off between computational demands and height estimation accuracy.

The authors acknowledge with thanks the logistic support by the project OP "Competitiveness" BG161PO0031-1.2.04-0044 financed by the Structural Funds of the European Union, and the project ISP1, financed by the Technical University of Varna, Bulgaria.

[G8.5]. I. Mporas, **T. Ganchev,** M. Siafarikas, N. Fakotakis (2007). "Comparison of Speech Features on the Speech Recognition Task", *Journal of Computer Science,* ISSN: 1549-3636, vol. 3, no. 8, 2007. 608-616.

In the present work we overview some recently proposed discrete Fourier transform (DFT)- and discrete wavelet packet transform (DWPT)-based speech parameterization methods and evaluate their performance on the speech recognition task. Specifically, in order to assess the practical value of these less studied speech parameterization methods, we evaluate them in a common experimental setup and compare their performance against traditional techniques, such as the Mel-frequency cepstral coefficients (MFCC) and perceptual linear predictive (PLP) cepstral coefficients, which presently dominate the speech recognition field. In particular, utilizing the well-established TIMIT speech corpus and employing the Sphinx-III speech recognizer, we present comparative results of eight different speech parameterization techniques.

This work was supported by the MoveOn project (IST-2005-034753).

[G8.6]. F. Feradov, **T. Ganchev,** I. Ivanov (2015). "Detection of Sleep Disorders Based on Time-Domain EEG Signal Descriptors". *Annals of TU-Varna* - 2015, pp.54-58.

Medical conditions related to sleep disorders influence the quality of life and may result to serious negative effects on the overall health of a person. The development of tools for the automated classification of sleep disorders is of significant importance as certain sleep disorders are symptoms of other serious diseases. In the present work, we study the appropriateness of three classification algorithms (MLP, SMO and J48) and evaluate their performance for a set of statistical signal descriptors on the task of Periodic Limb Movement Disorder (PLM) and REM Behavior Disorder (RBD) detection from electroencephalographic (EEG) signals.

The authors acknowledge with thanks the support received through the research projects NP8 “Development of advanced methods for automated analysis of EEG signals for detection of negative emotional states and neurological disorders”, SNP2 “Technological support for improving the quality of life of people with the Alzheimer disease” and – both financed by the Technical University of Varna, Bulgaria.

[G8.7]. N. Dukov, **T. Ganchev** (2017). "An empirical study on ReLU based neuron models of the LRPNN", *Proc. of the IEEE Biomedical Data Acquisition and Applications Workshop,* October 13-14, 2017, Technical University of Varna, Bulgaria. pp. 24-27.

In the present contribution, we explore different constructions of the Locally Recurrent Probabilistic Neural Network (LRPNN) neuron model. More specifically, we look at the activation function in the fourth layer of the neural network. Traditionally, a sigmoid activation function is used. In the current work, we show that classifying emotional states based on electroencephalographic (EEG) signals can be improved in terms of overall accuracy. This is achieved with the use of activation functions based on Rectified Linear Units (ReLU), instead of the sigmoid function. Moreover, with the improved results and simple computation the ReLU functions seem as a more appropriate choice when we consider accelerating the LRPNN.

The authors acknowledge with sincere thanks the support received through the research project entitled "NP7/2017 Study of Methods and Apparatus for the Acquisition of Biomedical Data in Mobile Setup", financed by the National Science Fund of Bulgaria and Technical University of Varna.

[G8.8]. Feradov F., **T. Ganchev,** N. Nikolov (2015). "A Study of Short-Time Energy as a Feature in BCI," *Proc. of the International Scientific Conference,* UNITECH-2015, November 21-22, 2015, Gabrovo, vol.2, pp.333-337.

The inherent complexity of Brain Computer Interfaces (BCI) restricts the opportunities for embedding them in assistance devices control and neuro-prostheses applications. In the present work we assess the feasibility of reduced complexity BCI which makes use of the short-time energy as a single signal descriptor computed over an ordinary multi-channel EEG signal. We consider classifying motor imagery from EEG and evaluate the mean recognition accuracy per class. The accuracy ranged between 65.6% and 71%. The experimental results indicate that the short-time energy is relevant feature and can be used in BCI, preferably in combination with other features.

The authors acknowledge with thanks the support received through the research projects PD8 "Development of advanced methods for automated analysis of EEG signals for detection of negative emotional states and neurological disorders" and SNP2 "Technological support for improving the quality of life of people with the Alzheimer disease" – both financed by the Technical University of Varna, Bulgaria.

[G8.9]. F. Feradov, J. Zhekov, and **T. Ganchev** (2015). "Assessing the Effectiveness of Time-Derived Statistical Descriptors in the Classification of Epileptic Seizures", Proceedings of the Innovation and Business Conference, I&B-2015 ‹‹Applied Technology for Health››, Oct. 9-10, TU-Varna

Epilepsy is a neurological disorder, which causes episodic loss of motor functions, loss of consciousness, and convulsions. Therefore, the prior detection and prediction of epileptic seizures is a topic of great social significance. The present work reports results of a comparative evaluation carried out for a number of statistical time-domain features of EEG signals, recorded on epileptic patients. These features are evaluated using three different classifiers – J48, SMO, and MLP. We report detection results in the range between 98.7% and 100%, in terms of mean classification accuracy.

The authors acknowledge with thanks the support received through the research projects PD8 "Development of advanced methods for automated analysis of EEG signals for detection of negative emotional states and neurological disorders" and SNP2 "Technological support for improving the quality of life of people with the Alzheimer disease" – both financed by the Technical University of Varna, Bulgaria.

[G8.10]. N. Dukov, **T. Ganchev**, and D. Kovachev (2015). "Reduced-complexity FPGA implementation of PNN for binary image classification". *Annals of TU-Varna* - 2015, pp.59-63.

In the present work, we study the feasibility of low-complexity implementation of PNN for classification of binary images. We evaluate the performance of software and FPGA-based hardware implementations of the PNN with and without the bias module. A comparative evaluation carried out on the cardiac Single Proton Emission Computed Tomography (SPECT) database demonstrates that significant reduction of the number of elements in the FPGA design can be achieved, without significant degradation of the classification accuracy. The reduced-complexity FPGA design allows efficient implementation of the PNN for classification of binary images with the Altera Cyclone IV chip.

The authors acknowledge with thanks the support received through the research projects PD5 “Study of biologically substantiated architectures of artificial neural networks for the identification of heart diseases and neurological disorders”, SNP2 “Technological support for improving quality of life of people with the Alzheimer disease”, and the NP4 “Capacity building for object-oriented FPGA design in support of knowledge-based economy” financed by the Technical University of Varna, Bulgaria.

[G8.11]. Feradov F., **T. Ganchev** (2014). "Detection of Negative Emotions from EEG Signals Using Time-Domain Features," *Proc. of the International Scientific Conference* UNITECH-2014, November 21-22, 2014, Gabrovo.

Electroencephalography provides the opportunity for direct monitoring and identification of emotions in the moment of their conceiving. Among the main challenges of automated emotion detection is the dependence on signal features that carry explicit unambiguous evidence about the emotions manifested while remaining insensitive to variability due to other brain activity. In the present paper, we report results from an experimental evaluation of the appropriateness of seven signal features derived directly from the time-domain EEG signal. The main advantage of these features is the low complexity of their computation. We report that the combined use of all seven features brings improvement of recognition accuracy, when compared to subsets studied in previous related work.

The authors acknowledge with thanks the financial and logistic support by the project OP "Competitiveness" BG161PO0031-1.2.04-0044 financed by the Structural Funds of the European Union, and by the project ISP1 financed by the Technical University of Varna, Bulgaria. The first author acknowledges with thanks the financial support by the OP "РЧР" BG051PO001-3.3.06-0005.

[G8.12]. F. Feradov, **T. Ganchev** (2017). "Ranking of statistical features of negative emotional states from EEG signals," *Proc. of the IEEE Biomedical Data Acquisition and Applications Workshop,* Oct. 13-14, 2017, Technical University of Varna, Bulgaria. pp. 20-23

In [G8.12], we present a study on the relevance of ten statistical features for the purposes of classification of negative emotional states. The features were evaluated using the ReliefFAttributeEval algorithm included in the WEKA machine learning toolbox. The evaluation was performed on EEG data taken from the DEAP database.

The authors acknowledge with sincere thanks the support received through the research project entitled "NP7/2017 Study of Methods and Apparatus for the Acquisition of Biomedical Data in Mobile Setup", financed by the Technical University of Varna.

[G8.13]. P. Zervas, **T. Ganchev** , and N. Fakotakis (2006). "Negative Emotional State Detection from Speech," *Proc. of the Proc. of Communication Systems Networks and Digital Signal Processing, CSNDSP-06, 19-21 July 2006. 310-313.*

This study reports results on the task of recognizing negative emotional states from speech with the employment of Probabilistic Neural Network and C4.5 tree classifiers. Four negative emotional states are considered: hot anger, cold anger, contempt, disgust, as well as neutral state. Appropriate speech data composed of short utterances extracted from the emotional database developed at the Linguistic Data Consortium at University of Pennsylvania, were used for training and testing the models. Each feature vector of the datasets was constituted of acoustical attributes of the signals related to energy and pitch. Results demonstrated that the Probabilistic Neural Network with an overall accuracy of 81.1 % outperformed the C4.5 approach, which attained 62.8 %.

[G8.14]. Dukov N., **Ganchev T.,** Kovachev D. (2014). "Empirical Study of PNN Sensitivity to the Exponential Function Approximation in FPGA Implementations", *Annals of TU-Varna* - 2014. vol.1, pp.53-58.

The hardware implementation of machine learning algorithms often requires the computation of nonlinear functions such as the exponential function, natural logarithm, cosines etc. In the present work, we study look-up table (LUT)-based approximations of the exponential function and the effects of their limited precision on the performance of a Probabilistic Neural Network (PNN)-based classifier. The empirical study was carried out on the cardiac Single Proton Emission Computed Tomography (SPECT) database, where we consider the task of two-class image classification for detection of myocardial perfusion.

The authors acknowledge with thanks the logistic support by the project OP "Competitiveness" BG161PO0031-1.2.04-0044 financed by the Structural Funds of the European Union, and by the project ISP1 financed by the Technical University of Varna, Bulgaria.

[G8.15]. A. Lazaridis, I. Mporas, **T. Ganchev** (2012). "Phone Duration Modeling of Affective Speech Using Vector Regression Support", *International Journal of Intelligent Systems and Applications* (IJISA), ISSN: 2074-904X (Print), ISSN: 2074-9058 (Online), vol.4, no.8, 2012. pp.1-9.

In speech synthesis, accurate modeling of prosody is important for producing high quality synthetic speech. One of the main aspects of prosody is phone duration. Robust phone duration modeling is a prerequisite for synthesizing emotional speech with natural sounding. In this work, ten phone duration models are evaluated. These models belong to well-known and widely used categories of algorithms, such as the decision trees, linear regression, lazy-learning algorithms and meta-learning algorithms. Furthermore, we investigate the effectiveness of Support Vector Regression (SVR) in phone duration modeling in the context of emotional speech. The evaluation of the eleven models is performed on a Modern Greek emotional speech database, which consists of four categories of emotional speech (anger, fear, joy, sadness) plus neutral speech. The experimental results demonstrated that the SVR-based modeling outperforms the other ten models across all the four emotion categories. Specifically, the SVR model achieved an average relative reduction of 8% in terms of root mean square error (RMSE) throughout all emotional categories.