

## **Резюмета на научните публикации**

на

доц. д-р инж. Тодор Димитров Ганчев, катедра „КНТ“  
представени за участие в конкурса за заемане на АД „Професор“  
в област „5 Технически науки“, по професионално направление  
„5.3 Комуникационна и компютърна техника“,  
учебна дисциплина “Изкуствен интелект”,  
обявен за нуждите на факултет ФИТА,  
катедра „Компютърни Науки и Технологии“,  
публикуван в ДВ брой 38 / 10.05.2019 год.

## **7 Резюмета на научните публикации представени за участие в конкурс за АД „Професор“**

### **7.1 Резюмета за публикациите от група „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.

В доклада [B4.1] е представен нов обобщен метод за обработка и анализ на физиологични сигнали, който в комбинация с класификатор с опорни вектори (SVM), позволяващ разпознаване и регистриране на негативни емоционални състояния, състояния с високи нива на емоционална възбуда и характерната комбинация от състояния с негативни емоции при висока степен на емоционална възбуда. Представени са методи за изчисляване на описатели на физиологични сигнали, в това число 16 описатели, изчислявани от електродермалната проводимост на кожата и 18 описатели, характеризиращи информацията, съдържаща се в ЕКГ сигналите.

Създаден е набор от записи на физиологични сигнали, събран специално за целта на настоящото изследване, разработката на модели и експерименталната валидация на обобщения метод. С помощта на този набор от записи, в доклада е изследвана работоспособността на обобщения метод при различни настройки на автоматичните детектори на негативни емоционални състояния, състояния на високо ниво на емоционално възбудждане и състояния, характерни с негативни емоции при високо ниво на емоционално възбудждане.

Полезнотта и работоспособността на обобщения метод е валидирана в контекста на интегрирана преносима система, която е в състояние да наблюдава непрекъснато определени физиологични сигнали на човешкото тяло, измежду които проводимостта на кожата и ЕКГ сигнали. Предложеният концептуален проект на персонална преносима система е изграден върху традиционната архитектура клиент-сървър. Събирането на физиологични сигнали се осъществява посредством интелигентни безжични сензори от типа Shimmer3 GSR+ и Shimmer3 ECG, които са свързани по Bluetooth канал към мобилен телефон. В мобилен телефон (смартфон) се изпълняват задачите, свързани с агрегиране на данните и предаване към сървъра. Сървърът изпълнява всички задачи, свързани с обработка и анализ на физиологичните сигнали. По такъв начин в сървъра се извършва цялата обработка, свързана с методите за обработка и анализ на данни и разпознаването на физиологични състояния, а в смартфона е реализиран потребителският интерфейс, който позволява на потребителя да следи и управлява процеса на събиране на данни и да визуализира резултатите от обработката.

В доклада са представени резултати за различни видове моделиране на гореспоменатите физиологични сигнали, в т.ч. с използването на знание за обобщената категория, към която принадлежат аудио-визуалните стимули, подавани към участниците и само информация, получена от самооценката на въздействията от участниците. Експерименталните резултати от изследванията с базовите модели, показват точност на разпознаване с горна граница до 85%, което подсказва добър потенциал за подобреие при реализирането на необходимата за работата на системата функционалност.

**[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.

В доклада [B4.2] е представен нов метод за автоматичен ревизиран подбор на описатели, при който процесът на рафиниране на първоначалния пълен набор от описатели, преминава през три основни етапа. Идеята на метода е при подбора на описатели, да се надгради методът за оценка на характеристиките на описателите чрез критерия за разделяне на Фишер. Това се осъществява посредством последващи обработки, включващи (i) допълнителен подбор, използващ предварително знание за проблема и (ii) допълнително рафиниране на разнообразието в набора описатели, с цел намаляване изменчивостта в окончателния набор от избрани описатели. При допълнителното рафиниране на разнообразието се налагат ограничения при избора на описатели, въз основа на знания, специфични за конкретната задача, като това се очаква да улесни моделирането и класификацията на следващите етапи при разпознаване на определени физиологични състояния.

Предложеният метод е валидиран чрез експериментална постановка, ориентирана към откриване на състояния на остьр стрес на базата на физиологични сигнали. За целта е използвана базата данни с физиологични записи ASCERTAIN, от която са взети записи на ЕКГ сигнали и повърхностно съпротивление на кожата за 58 человека. От тези физиологични сигнали са изчислени 43 описатели, от които чрез специално създадена процедура за подбор, са избрани основните 18, за които се знае, че са релевантни към решаваната задача, т.е. тези се основават на предварително знание за проблема.

Експерименталните резултати потвърждават, че предложеният метод за ограничен подбор на описателите постига предимство, в сравнение с три други случая:

- (i) използване на пълния набор от описатели,
- (ii) използване на подмножество, избрано въз основа на предварително познание за проблема, и
- (iii) използване на подмножество, избрано единствено въз основа на разделянето на критерия на Фишер.

В сравнение с тези базови методи е постигнато относително усреднено подобрение на точността на разпознаване на състояния на остьр стрес, със средно 5.4% (за всичките 58 человека от ASCERTAIN).

Основно достойнство на новия метод е, че предлага механизъм за насочване и донастройка на процеса на избор на описатели, което позволява избраният набор от описатели да отчита релевантността на описателите за решаване на конкретната задача и в същото време, да се отчитат индивидуалните физиологични характеристики на всеки човек. Използването на предложения нов метод води до избор на подмножества от описатели, които се припокриват частично и са релевантни към съответната експериментална постановка на задачата, но отчитат особеностите на всеки човек.

**[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

В доклада [B4.3] е представен нов триетапен метод за избор на описатели, който се основава на използването както на обобщена, независеща от индивида оценка на атрибутите, така и на оценка на описателите на сигнали, отчитаща индивидуалната особеност за всеки отделен човек. Първите две стъпки от метода имат за цел да изберат подмножество от атрибути, които са универсални и общи за голяма популация от индивиди, като това се извършва чрез селекция на описатели, които се избират многократно за големи групи от потребители. В последствие, тази селекция се рафинира чрез ограничаване на избора до подгрупа от описатели, специфична за определен човек, която е получена след прилагане на критерия за разделяне на Фишер. В резултат на сечението на тези подмножества се получава ново подмножество от атрибути, което е едновременно специфично за така формулираната задача и същевременно персонализирано към конкретен човек, според качеството на регистрираните от него сигнали.

Предложеният нов метод е валидиран с помощта на базата данни от физиологични сигнали ASCERTAIN, от която са използвани записите на ЕКГ сигнали и повърхностно съпротивление на кожата за всичките 58 человека. Експерименталната постановка е насочена към разпознаване на физиологични състояния, асоциирани със стрес, които се характеризират с високо ниво на емоционална възбуда при негативни емоции. Изследвани са голям брой различни настройки на метода и е извършен анализ на получените подмножества от описатели. Проведено е сравнително експериментално изследване на точността на класификация за осем набора от описатели, в това число за

- (i) три различни настройки на предложения метод,
- (ii) три различни настройки на метода за селекция на най-често избираните описатели след селекция с метода на Фишер,
- (iii) индивидуалните подмножества от описатели, които са избрани чрез метода на Фишер поотделно за всеки човек,
- (iv) пълния набор от 43 описатели.

Експерименталните резултати са в подкрепа на тезата, че предложеният метод предлага предимства по отношение на точността на откриване на физиологични състояния, характеризиращи се с високо ниво на емоционална възбуда при негативни емоции, в сравнение с други стратегии за подбор на подмножества от описатели. В резултат на прилагането на новия метод за подбор на описатели е отчетено подобреие на относителната точност на разпознаване с над 5.3 % спрямо пълния набор от описатели.

**[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

В доклада [B4.4] се изследват два различни подхода за автоматизирано откриване на негативни емоционални състояния, както и на състояния, характеризиращи се с висока емоционална възбуда и негативни емоции (HANV), които са близки до острия стрес в специфичен контекст. Независимо от крайната цел, тези два подхода използват алтернативни формулировки на задачата за определяне на съответното състояние – чрез използване на тагове, които са:

- 1) До известна степен обективни, т.к. са общи за всички хора и са дефинирани единствено от съдържанието на всеки конкретен стимул (повече или по-малко консенсусно за голяма група хора), или
- 2) Субективни и индивидуални за всеки потенциален ползвател на тази технология, т.к. се извеждат въз основа на само-определението на всеки човек за ефекта от прилагането на конкретен стимул.

Използвайки тези две формулировки на задачата, в доклада се изследват възможностите за създаване на автоматични детектори на HANV състояния и на негативни емоционални състояния. Изследвани са последствията и сложността на задачата при двете формулировки и възможностите за автоматизирано разпознаване на тези състояния. За целта е проведено експериментално оценяване на предимства на моделирането на базата на независими тагове, като се използват етикетите присъединени към аудио-видео стимулите от създателите на базата данни, както и тези получени от самооценката на конкретните хора.

За целите на експерименталното изследване е създадена автоматизирана система за разпознаване на емоционални състояния, въз основа на физиологични сигнали, в т.ч. ЕКГ и повърхностно съпротивление на кожата. В доклада са разгледани общата архитектура на системата, както и методите, използвани на изчисляване на описателите на физиологични сигнали, методите за допълнителна обработка и селекция на описателите, като и метода за класификация на данните при откриването на (i) негативни емоционални състояния и (ii) HANV състояния.

Анализът на резултатите от изследването показва, че по-точно и надеждно разпознаване на състоянията, характеризиращи се с висока емоционална възбуда и негативни емоции, се получава, когато се разчита на субективните тагове, които са индивидуални за всеки потенциален ползвател. В тази връзка е показано, че използването на субективните тагове, определянето на които изисква малко допълнително усилия, позволява да се постигне относително подобряване на точността на откриване на HANV с до 5%.

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

Изследванията, свързани със създаване на методи за автоматично разпознаване на емоционални състояния и оценка на нивата на стрес, до голяма степен зависят от наличието на подходящи ресурси, в това число на бази данни със записи на физиологични сигнали, получени след прилагането на определени видове стимули.

В доклада [B4.5] е представен цялостният дизайн на система за събиране на базата данни, SLADE, която е създадена специално в помощ на научноизследователските дейности по разпознаване на емоционални състояния и оценка на нивата на стрес. SLADE има за цел да улесни разработването на автоматизирани инструменти и услуги за оценка и наблюдение на нивата на стрес. Накратко базата данни се състои от синхронизирани записи на 16-канални ЕЕГ сигнали, 3-канални ЕКГ сигнали, повърхностната температурата на кожата (ST) и галваничната реакция на кожата (GSR). Подробно са описани отделните модули и компоненти, съставляващи стенда за сбор на данни, както и специално създаденото програмно осигуряване, с помощта на което се извършва синхронното изльчване на аудио-видео стимулите и събирането на записите на физиологични сигнали.

Освен представянето на количествените и качествени параметри на системата и на базата данни от физиологични данни, е извършена експериментална работа за валидиране полезността на базата данни SLADE. Представени са експериментални резултати от специално създадени за целта базови системи за разпознаване на (i) емоционална възбуда, (ii) негативни емоционални състояния, както и (iii) състояния, характеризиращи се с висока степен на емоционална възбуда при негативни емоционални състояния. Последните са асоциирани с групата състояния, имащи отношения към острая стрес.

В доклада подробно е описана архитектурата на системата за разпознаване на емоционални състояния, извличането на описателите, допълнителните обработки на описателите и процеса на класификация, като са представени експериментални резултати, показващи висока степен на точност при комбинираното използване на голям брой описатели, извлечени от синхронизирани записи на 16-канални ЕЕГ сигнали, 3-канални ЕКГ сигнали, повърхностната температурата на кожата (ST) и галваничната реакция на кожата (GSR).

Научноизследователската и научно-приложна дейност, представени в доклада, са частично подкрепени от проект НП7/2017 “Методи и средства за събиране на биомедицинска информация чрез мобилни информационни системи“, финансиран от държавния бюджет и Технически университет- Варна.

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

В доклада [B4.6] се разглеждат качествата на линейните кепстрални коефициенти като описатели на ЕЕГ сигнала, използвани за нуждите на автоматично разпознаване на негативни емоционални състояния. Изследваните описатели са изчислени, използвайки три различни метода за предварителна обработка на ЕЕГ сигналите:

- (i) след сегментиране на сигнала с предварително подбрана прозоречна функция, се изчислява отделен набор от описатели за всеки отрязък от сигнала,
- (ii) след сегментиране на сигнала и усредняване на отрязъците, като поелементно средно аритметично на отчетите, се изчислява само един набор от описатели за осреднения отрязък,
- (iii) директно изчисляване на един набор от описатели за целия запис.

При всеки от тези методи описателите на ЕЕГ сигнала са изчислени за 6 различни разделения на честотната лента, които се реализират с 10, 15, 20, 30, 45 и 60 филтъра.

Сравнителният анализ на различните конфигурации в тези три метода се извършва чрез експериментален протокол, използващ базата данни от физиологични сигнали DEAP, от които са извлечени само ЕЕГ записите. Данните на участниците се разделят на две групи, в зависимост от оценката по критерия „харесване“, поставена от участниците. Записите, получили оценка по-ниска от 4, се маркират като „негативни“, докато записи с оценка по-висока от 4, се маркират като „други“ (неутрални и позитивни). При провеждането на тестовете се използват данните на 24 участници, при които след разпределението на записи всеки от двата класа съдържа не по-малко от 20% от общия обем ЕЕГ данни. Полезнотта на отделните набори описатели е оценена чрез автоматизирана система за детекция на негативни емоционални състояния, използваща класификационния алгоритъм C4.5 в работната среда WEKA.

Получената средна точност при класификация варира в рамките 70.7% – 86.9%. Анализът на получените резултати показва, че точността на линейните кепстрални коефициенти и спектралната логаритмична енергия е много близка в случаите, при които се ЕЕГ сигналите се сегментират посредством прозоречна функция, като в тези случаи се наблюдават и най-високите постигнати точности. И при двете категории описатели, при прилагането на усредняване на ЕЕГ отрязъците, се наблюдават сравнително ниски резултати.

Сходни резултати се наблюдават и при използването на описатели, изчислени от несегментиран ЕЕГ сигнал, посредством банки с малък брой филтри – 10 и 15 филтъра. При увеличаването на броя филтри обаче се наблюдава и повишаване на точността на класификация, като тази зависимост е по-ясно изразена при използването на спектралната логаритмична енергия, където разликата в средната точност на класификация между използването на 10 и на 60 филтъра е 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.

Изследванията, представени в доклада [B4.7], са насочени към изучаването на пространствените зависимости между ЕЕГ сигналите и създаването на описатели, които да отчитат такива пространствени корелации. В доклада е предложен нов описател на ЕЕГ сигнала, представящ *декорелираната енергия на време-пространственото разпределение на ЕЕГ активността* и е изследвана приложимостта му за разпознаване на негативни емоционални състояния. Представена е експериментална оценка на точността при класификация на негативни емоционални състояния. Експерименталната постановка е организирана чрез базата данни DEAP и групиране на записите по два различни критерия – „харесване“ и „валентност“.

Процесът на извлечане на предложения в доклада нов описател се състои от четири стъпки:

- 1) преобразуване на оригиналния многоканален ЕЕГ сигнал във време-пространствено представяне на сигнала. Преобразуването се осъществява посредством прилагането на кратковременно преобразуване на Фурье върху пространственото измерение на сигнала, т.е. едновременно по всички канали. Това се постига, като кратковременното преобразуване на Фурье се прилага върху отчетите, снети от всички канали в даден момент от време.
- 2) получената спектrograma на ЕЕГ сигнала се мащабира посредством набор от припокриващи се триъгълни прозоречни функции.
- 3) изчислява се логаритмичната енергия на пространствената активност за избраният период от ЕЕГ сигнала. Вземайки изчислената време-пространствена енергия за всеки един прозорец, получаваме представяне на спектrogramата мащабирано по време.
- 4) прилага се дискретно косинусово преобразуване. По този начин енергията, изчислена за всеки един времеви прозорец, се декорелира и получаваме декорелираните време-пространствени коефициенти (Decorrelated Spatio-Temporal Coefficients - DSTC).

Количествена оценка за полезността на описателите се извършва чрез експериментален протокол, използващ ЕЕГ записи от базата данни DEAP и автоматизиран детектор на негативни емоционални състояния. Изготвяните модели са индивидуални за всеки участник.

Получената средна точност варира в рамките 70.8% - 78%. Анализът на експерименталните резултати води до заключението, че чрез предложението нов тип описател се постига висока класификационна точност, която е съпоставима с тази на други утвърдени описатели. Следователно декорелираната енергия на време-пространственото разпределение на ЕЕГ активността може да бъде успешно използван при решаването на задачи, свързани с класификацията на негативни емоционални състояния.

**[B4.8].** T. Kostoulas, T. Winkler, **T. Ganchev**, N. Fakotakis, J. Köhler (2013). “The MoveOn Database: 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**)

В статията [B4.8] е представен цялостният дизайн и реализацията на аудио база данни MoveOn, съдържаща записи на стресирана реч от мотоциклистни полицаи, записи на съществуващи акустични събития от други източници и шумове, от движението на мотоциклет по градската и извънградска пътна мрежа. Характерно за записаната реч е, че тя е представителна за типичната комуникация между полицейските служби за управление на мисии и полицаи -мотоциклисти, работещи в условия, характеризиращи се с комбинация от физически и когнитивен стрес и високо ниво на шум и смущения от околната среда. Базата данни е целенасочено проектирана в подкрепа на изследванията по създаване на речеви диалогови информационни системи за полицията в Обединеното кралство, като целевата група са полицаи-мотоциклисти от специализирано подразделение за оперативна дейност и преследване на заподозрени лица.

В статията са представени основните предизвикателства, свързани с избора на оборудване, организиране на сесиите за запис и с описани решенията и компромисите, които са наложени от трудности, свързани с естеството на дейността на мотоциклистите. Уникалните характеристики на базата данни MoveOn са в резултат от изисквания на специфичната приложна област - информационна поддръжка и оперативна система за командване и контрол на полицейските сили на мотоциклети и също от спецификата на неблагоприятната акустична среда при движение по пътната мрежа, а когато се налага и извън нея.

В статията е представено подробно статистическо описание на базата данни, включително и (i) анотациите на речта, (ii) емоционалната окраска и нивото на стрес в речта, както и (iii) типичните категории акустични събития и (iv) типичните категории шумове от естествената околната среда. Представено е разделяне на данните в подмножества, необходими за създаването на акустични модели и за тяхната експериментална оценка.

Анализът на резултатите от проведените експериментални изследвания на прецизността и точността при автоматично разпознаване на реч в условия, характеризиращи се с комбинация от физически и когнитивен стрес, илюстрират степента на сложност на конкретната оперативната среда, в която работят полицайите от мотоциклистното подразделение.

Изследванията, представени в статията, са финансиирани по проекта MoveOn (IST-2005-034753), който е съфинансиран по програмата FP6 на Европейската комисия.

[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.

В статията [B4.9] е представена нова архитектура на речеви интерфейс, в който е осъществена функционалност за разпознаване на емоционална реч и емоционални състояния като гняв, отегчение и др. Този интерфейс е специално проектиран за интегриране в платформата PlayMancer, която е предназначена за създаване на сериозни игри, подпомагащи когнитивното лечение на пациенти с поведенчески отклонения и психични разстройства. Използването на компютърни игри подпомага пациентите да развиват нови умения за самоконтрол и справяне с негативни емоции при ежедневни стресови ситуации.

Предложената архитектура на речевия интерфейс надгражда платформата Olympus/RavenClaw, към която са добавени нови компоненти за разпознаване на определени емоционални състояния от реч, биосигнали, камера и са осъществени допълнителни разширения за нуждите на разработчиците на сериозни игри.

За валидиране работата на новосъздадения речеви интерфейс е използван специално създаден за целта набор от записи на емоционална реч, специфични за конкретната приложна област. Проведените експериментални изследвания показват, че използването на емоционалната реч влошава прецизността на речевия интерфейс, което налага разработването на специални акустични модели за разпознаване на емоционална реч.

Като цяло резултатите от разпознаване на реч и откриване на емоционалните състояния, са добри. Детекторите на емоционални състояния като интерес, гняв и скука, показват задоволителна точност и прецизност и допринасят за по-успешното моделиране на емоционален статус на пациента.

Изследванията, представени в статията, са финансираны по проекта PlayMancer (FP7-ICT-215839-2007), който е съфинансиран по програмата FP7 на Европейската комисия.

**[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, 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**)

Статията [B4.10] представя концепцията и основните резултати от изпълнението на новаторския и високо оценен европейски изследователски проект PlayMancer, който проектира, разработи и оцени приложимостта на новосъздадената сериозна видео игра, предназначена за подпомагане възстановяването при пациенти с нарушен контрол на импулсивността, страдащи от нарушенна ментална нагласа, поведенчески и емоционални разстройства и други свързани заболявания. Видеоиграта използва *биофийдбек* за подпомагане пациентите да усвоят умения за релаксация, да усвоят стратегии за самоконтрол и да си изработят нови стратегии за емоционално регулиране. В статията е представено кратко описание на разработената видеоигра, обосновка, потребителски изисквания, иновативната технология за много-модално разпознаване на емоционални състояния и предварителни резултати, относно полезнотта при лечението на определени психични разстройства (пристрастеност към хазарта, хранителни разстройства). Накратко, в рамките на проведените изследвания е показано, че създадената видеоигра има потенциал и капацитет за промяна на основните нагласи, поведението и емоционални процеси при пациенти с импулсивни нарушения. Чрез играта се предоставят възможности за нови режими на взаимодействие и обективна оценка на емоционалното състояние на пациента. Това се дължи главно на реализираните нови функционалности, използващи иновативни методи за разпознаване на емоции от реч, изражението на лицето и физиологични сигнали.

Въз основа на база данни от реч, видео и физиологични сигнали (в т.ч. ЕКГ, повърхностно съпротивление на кожата, честота на дишане и др.), специално създадена за нуждите на това изследване, са създадени статистически модели на емоционални състояния, както и нови методи за много-модално разпознаване на емоции и специфични ментални състояния. Тези иновативни методи за много-модално разпознаване на емоционални състояния, са интегрирани във видеоиграта, с цел създаване на *биофийдбек*, което позволява логиката и стратегията на развитие на сценариите, да зависят от моментното емоционално и психологическо състояние на играещия. Така се постига пряка връзка между състоянието на играещия в реалния свят и поведението на екранния герой в заститения свят на играта. По такъв начин, използването на *биофийдбек* създава условия за подпомагане придобиването на умения за релаксация, самоконтрол и емоционално регулиране.

Макар и да съществуват предишни проучвания, показващи полезнотта на сериозните видеоигри при лечението на някои аномални поведения, или като допълнителни средства при лечението на заболявания в няколко области (шизофрения, астма или двигателна рехабилитация), тази иновативна методологична и технологична разработка запълва липсата на изследвания и възможността специално предназначени за целта сериозни игри, да се използват за подпомагане лечението на психични разстройства.

Изследванията, представени в статията, са финансираны по проекта PlayMancer (FP7-ICT-215839-2007), който е съфинансиран по програмата FP7 на Европейската комисия.

**[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

В доклада [B4.11] са изследвани възможностите за подобряване на точността при разпознаване на пет категории емоции от спонтанна реч, независимо от личността на говорещия, чрез подбор на релевантен набор от описатели на речевия сигнал. Отправна точка на изследването е набора от 384 описатели на речта, дефинирани от организаторите на състезанието Interspeech 2009 Emotion Challenge. За нуждите на конкретното изследване е създадена система за разпознаване на пет емоционални състояния от спонтанна реч. Описателите на речевия сигнал се изчисляват съгласно методите, дефинирани от организаторите на състезанието, а разпознаването се осъществява чрез UBM-GMM метод за моделиране на категориите и класификация. В случая UBM-GMM метода, въпреки че е често използван при разпознаване на диктори и в изследванията по разпознаване на реч, не е добре изучен в приложения за разпознаване на емоционални състояния от спонтанна реч, т.е. този аспект се явява научно-приложна новост в настоящото изследване.

Основният принос на представеното изследване се състои в систематичната оценка на относителната значимост на отделните описатели на речта и тяхната релевантност към задачата за разпознаване на емоционални състояния. За целта базовият набор от 384 описатели е подложен на систематична оценка с метода за търсене BestFirst при десетократна оценка на качеството на отделните описатели. За целите на изследването са използвани 10 различни подмножества на набора от развойни данни. След десетократната оценка за релевантност са подбрани 56 от базовите описатели на речта, включвайки само тези, които са селектирани най-малко 5 пъти в десетте теста. Тези 56 описатели са сортирани въз основа на оценката на важността им, след което са съставени подмножества, включващи само първия, първите два, първите три, и т.н. Проведени са общо 56 експеримента за оценяване точността на разпознаване на петте категории емоции, за всяко подмножество. Въз основа на анализ на експерименталните резултати е избран набор, включващ само първите 49 описатели от 56, т.к. добавянето на следващите не допринася за получаването на по-висока точност.

Експерименталната постановка следва протокола, дефиниран от Interspeech 2009 Emotion Challenge. За целта на изследването е използвана базата данни от спонтанна емоционално-оценетена реч FAU Aibo Emotion Corpus, съдържаща записи реч от 51 деца докато играят с обучаваш робот Aibo, командвайки робота да извършва предварително дефинирани последователности от действия, които са предназначени да провокират емоционални реакции.

Анализът на резултатите показва, че предложеният метод за подбор на набор от описатели на речта допринася за подобряване на точността при разпознаване на емоционални състояния спрямо тази, получена чрез базовия набор. Същевременно се намалява необходимата изчислителна мощ и времето за обработка на данните.

Изследванията представени в доклада са финансираны по проект PlayMancer (FP7-ICT-215839), който е частично финансиран по програмата FP7 на Европейската комисия.

**[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.

Интерфейсът човек-компютър (HCI), особено в областта на игрите, цели да имитира във възможно най-голяма степен естественото общуване човек-човек, което по своята същност се явява много-модално - включва говор, зрение, докосване, обоняние, вкус. По тази причина в областта на сериозните игри все по-често се използват допълнителни модалности, например като биосензори за физиологични сигнали, оборудване за проследяване на движението и др. В този контекст създаването на много-модални сериозни игри е сложен и скъп процес с дълъг цикъл на развитие, което допълнително затруднява създаването на нови жанрове и въвеждането на иновации в игрите.

В доклада [B4.12] са представени концептуалният проект и основните функционалности на много-модалния диалогов мениджър, който осъществява обработката и интерпретацията на информацията и цялостното управление на интерфейса човек-машина в платформата PlayMancer. Представени са общата архитектура и работата на отделните компоненти, като основните акценти са на разширена функционалност, дължаща се на нововъведени компоненти за разпознаване на емоции от реч и аудио, видео и физиологични сигнали, за откриване и проследяване на траекторията на движение, както и нови инструменти за конфигуриране на речевия диалог и цялостното управление и конфигурация на много-модалния интерфейс.

Основни моменти в доклада са дефиниране на архитектурата на речевия интерфейс и изясняване функционалностите, с които отделните компоненти допринасят за осъществяване на гласово общуване човек-машина, включващо някои елементи, присъщи за естествения диалог човек-човек, в т.ч. разпознаване на емоционални състояния.

За по-голяма точност и достоверност при откриването на емоционални състояния е създаден компонент за синхронизиране и обединяване на данните от различни модалности, където се осъществява комбиниране на информация, получена от няколко информационни потока, в конкретния случай включващи данни от биосензори (ЕКГ, съпротивление на кожата, честота на дишане) и оценка на емоционалността на речта. От своя страна информацията за емоционалното състояние, се използва при оценяване на когнитивното усилие на играещия, с цел да се поддържа определена степен на трудност и динамичност на играта, гарантираща ангажираността и мотивацията на пациента.

Изследванията, представени в доклада, са финансиирани по проект PlayMancer (FP7-ICT-215839), който е частично финансиран от Европейската комисия в рамките на Седма рамкова програма.

**[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

В доклада [B4.13] е предложена общата архитектура на интелигентна диалогова система, възприемаща спонтанна реч за управление на домашни електроуреди в среда на интелигентен дом. Интелигентността в предложената архитектура произлиза от заложената функционалност за автоматична идентификация на говорещия и способността за автоматично разпознаване на емоциите на говорещия от речевите сигнали.

Представен е нов метод за разпознаване на негативни емоционални състояния от спонтанна реч, която е присъща за общуване в среда на интелигентен дом. Методът използва модел на емоциите, създаден въз основа на два набора от данни – първия съдържа реч с актьорска интерпретация на емоции, а втората се състои от спонтанна реч, записана в среда на интелигентен дом. Извличането на описатели от речта следва широко използвани методи за оценяване периода на основния тон, за оценяване на относителния дял на хармонични към шумови компоненти в речта, за оценката на кратковременната енергия на сигнала и за изчисляване на 13 кепстрални коефициенти след деформиране на честотната лента, съгласно Мел скалата (MFCC). Моделирането на емоциите се извършва посредством смес от Гаусови функции (GMM). На базата на предложения метод е създаден детектор на негативни емоционални състояния (DNES), който служи за усъвършенстване на потребителския речеви интерфейс на диалоговата система на интелигентен дом. С помощта на DNES диалоговата система може да разбере, дали потребителят изпитва негативни емоции заради затруднения да постигне желана цел при управлението на домашната среда и електроуреди, или по друга причина. Чувствителността на системата към емоционалната окраска на речта позволява при нужда да предложи помош на потребителя, и/или да се адаптират настройките на потребителския интерфейс, и/или да се адаптира стратегията за водене на диалога с потребителя, за постигане на по-успешно взаимодействие между човека и компютъра, командващ интелигентния дом.

В доклада са представени резултати от изчерпателен анализ на работата на предложения метод при различни видове входна реч и е показана точността на разпознаване на негативни емоции за множество работни точки, дефиниращи различна толерантност към пропуск при разпознаване на негативни емоции от първа проява. Получените експериментални резултати показват, че точността за разпознаване на негативните емоции в спонтанна реч, която се използва в реални условия, е по-ниска от съобщаваните в литературата, при използването на бази данни с актьорска интерпретация на емоционална реч. В заключение, би следвало да се отбележи, че получените резултати за точността при автоматично разпознаване на емоции от спонтанна реч се възприемат като обещаващи, особено като се има предвид, че над 20% от хората често не могат да разграничват емоцията на други хора, съдейки само от кратък сегмент от реч.

Изследванията, представени в доклада, са частично подпомогнати от проект LOGOS (ЕНГ-102): “A general architecture for speech recognition and user friendly dialogue interaction for advanced commercial applications”, финансиран от Генералния секретариат за Научни изследвания и технологии (GSRT) на Гърция.

## 7.2 Резюмета на публикациите от група „Г.7

[Г7.1]. N. Dukov, T. Ganchev, D. Kovachev (2017). „FPGA implementation of the Locally Recurrent Probabilistic Neural Network“. Proc. of the IIITI-2017, September 14-16, 2017, Varna, Bulgaria, in “*Advances in Intelligent Systems and Computing*”, ISSN: 2194-5357, vol. 680, pp. 419-428.

В доклада [Г7.1] се представят (i) аппаратна реализация на локално рекурентен слой на Локално рекурентната вероятностна невронна мрежа (LRPNN) в FPGA чип и (ii) изследване на нейната работоспособност. Локално рекурентната вероятностна невронна мрежа е изградена от входен слой, три скрити слоя и изходен слой. Първите два скрити слоеве наподобяват оригиналната вероятностна невронна мрежа (PNN), докато третият скрит слой, наречен локално рекурентен слой е способен да моделира корелации във времеви последователности от наблюдения. Важен аспект при проектирането на хардуерния дизайн поради спецификата на LRPNN архитектурата е използването на модули с много висока двоична разрядност при представяне на стойностите на коефициентите на невронната мрежа. Въпреки че е скъпо от гледна точка на наличните ресурси в FPGA чипа, това е необходимо, за да се редуцира натрупването на добавената грешка от квантуване по ниво поради многобройните обратни връзки от невроните в рекурентния слой.

За изследване на работоспособността и прецизността на предложения дизайн е дефиниран експериментален протокол за разпознаване на негативни емоционални състояния от ЕЕГ сигнали. За целта е използвана базата данни от физиологични сигнали DEAP, от която са извлечени записите на ЕЕГ сигнали от 32-канален електроенцефалограф за 10 человека. От ЕЕГ сигнала са изчислени описатели, представляващи краткосрочна енергия на сигнала за преминаващи отрязъци от сигнала. Описателите, изчислени от всичките 32 канала за конкретен момент от времето, се обединяват в общ входен вектор за невронната мрежа. Последователностите от входни вектори позволяват да се моделира динамиката на изменение на енергията на сигнала във времето и пространството (според позициите на отделните канали).

В доклада се използват тегловни коефициенти на локално рекурентния слой на LRPNN, които са изчислени в среда на MATLAB с 64-битова точност с плаваща запетая, след което са пренесени в хардуерния дизайн със съответното закръгляне. Поотделно е изследвана работоспособността на аппаратна реализация на локално рекурентен слой на LRPNN и е извършен сравнителния анализ със софтуерната реализация в среда на MATLAB с неограничена и с ограничена точност на представяне на стойностите.

Анализ на резултатите от експерименталната валидация на аппаратната реализация на локално рекурентния слой показва, че има приблизително 1% процентна разлика между точността, постигната с помощта софтуерната реализация в среда на MATLAB (реализация с 64-бит CPU) и хардуерната реализация в FPGA чип.

Изследването представено в доклада е частично подкрепено от проекта НП7/2017 „Изследване на възможности за проектиране и хардуерна реализация на монолитни интегрални схеми и електронни системи с програмируеми аналогови матрици“, целево финансиран от Технически университет – Варна.

- [Г7.2]. V. Markova, K. Kalinkov, P. Stanev, T. Ganchev (2017). “Automated Stress Level Monitoring in Mobile Setup”. *Proc. of the ITI-2017*, September 14-16, 2017, Varna, Bulgaria, in “Advances in Intelligent Systems and Computing”, ISSN: 2194-5357, vol.680, pp.323-331.

В доклада [Г7.2] се представя концептуалният проект и дизайна на отделните модули на мобилна система за оценка на нивата на стрес в реално време. Системата комбинира информацията от интелигентни безжични сензори, които се носят върху облеклото или тялото и *Cloud computing* технология за събиране, обработка и анализ данните. За целта системата регистрира физиологични сигнали, като повърхностно галванично съпротивление на кожата (GSR) и повърхностна температура на кожата. В доклада е изследван конкретен работен сценарий, при който в реално време се събират, натрупват и анализират данни за конкретен човек.

Акцентът на изследването е експериментална валидация на концептуалния проект. Целта на изследването е да се провери до каква степен грешките и смущенията, натрупващи се в процеса на събиране, пренасяне и обработка на данните, биха повлияли възможността да се регистрира емоционална възбуда или емоционалното състояние на човека в мобилна конфигурация.

Анализът на експерименталните резултати показва подкрепа за хипотезата, че е осъществима работоспособна реализация на заложената функционалност, въз основа на описателите изчислени от GSR и показанията на температурата на кожата -- и в частност, може да се получи качествена оценка на емоционална възбуда в мобилна конфигурация. Такава функционалност улеснява развитието на технологиите за мониторинг на стреса в реално време в контекста на мобилни приложения.

В допълнение, в доклада се оценява възможността за откриване на отрицателни емоционални състояния въз основа на същия набор от описатели на физиологичните сигнали. Макар че наблюдаваната точност на разпознаване на емоционалната валентност не съответства на изискванията на приложенията от реалния живот, експерименталните резултати се считат за интересни, тъй като те показват възможността за откриване на специфични случаи – състояния, характеризиращи се с висока степен на емоционална възбуда при ясно изразени негативни емоции (HANV).

В дискусията към доклада е подсказано, че в специфични ситуации и контекст, събитията от типа HANV могат да се интерпретират като начало на остро стресово състояние, при което на лицето може да се подскаже как да продължи мисията си. Хипотетично, последното може да бъде полезно за подобряване на управлението на стреса във високорискови професии, в подкрепа на екипи, участващи в спешна медицинска помощ и в рехабилитацията на пациенти, които се възстановяват от тежка травма, като наранявания на гръбначния мозък, долните крайници и др.

Научноизследователската и научно-приложна дейност, представени в доклада са частично подкрепени от проект НП7/2017 “Методи и средства за събиране на биомедицинска информация чрез мобилни информационни системи“, финансиран от държавния бюджет и Технически университет- Варна.

[Г7.3]. S. Ntalampiras, D. Arsić, 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)

В статията [Г7.3] е представен концептуалния проект и резултатите от създаването на много-модалната и много-сензорна научноизследователска база данни PROMETHEUS. Тази база данни представлява основен ресурс, създаден в подкрепа на изследователската и развойна дейност по тематиките, свързани с наблюдение и тълкуване на човешкото поведение в неограничена вътрешна и външна среда. Базата данни PROMETHEUS е уникална, не само заради иновативния набор от сензори използвани в различните сценарии, но също така и заради насочеността на дизайна, които обхваща редица приложения от реалния свят, свързани с автоматизацията на интелигентни системи за домашно и външно наблюдение на обществени пространства, в т.ч. големи търговски комплекси, летища, улици и площици, отворени градски пространства.

Накратко, базата данни съдържа повече от 4 часа синхронизирани записи от хетерогенни калибрирани сензори (в т.ч. инфрачервен сензор за откриване на движение, термовизионни камери, видеокамери за преглед / наблюдение, видеокамери с близък поглед, 3D камера, стереоскопичен фотоапарат, общ предназначени за видеокамера, микрофонни решетки и оборудване за улавяне на движение), събрани в сценарии, наподобяващи средата, реалистични дейности и ситуации, в интелигентен дом, летища, публични пространства, банкомати и др.

С помощта на опитни актьори, статисти и допълнителен персонал бяха реализирани многобройни сценарии, включващи дейности на един човек и/или взаимодействия между няколко човека, както и сценарии, осъществяващи взаимодействия между групи от хора. В тези сценарии актьорите, статистите и допълнителният персонал бяха инструктирани да проиграват ситуации, свързани с редица типични и нетипични поведения, както и симулации на кризисни и аварийни ситуации.

За нуждите на научноизследователските и развойни дейности по откриване и анонимно проследяване на хора и моделиране и разпознаване на човешкото поведение в конкретната среда, в това число антисоциални, агресивни поведения, незаконосъобразни действия и вандализъм в публични пространства, за избрани сцени от базата данни са направени анотации. Цялата аудио част на базата данни е анотирана за нуждите на откриването и разпознаването на звукови събития, откриване и разграничаване на източника на звук, разпознаването на емоции и др.

Възможностите за научноизследователска работа с базата данни са демонстрирани чрез експерименти по откриване, разграничаване и идентификация на човешка фигура, нейната траектория на движение и действия, посредством комбиниране на информация, получена от различни видове оптични сензори и камери. Също така са представени резултати по разпознаване на звукови събития в затворени пространства и на открито, в т.ч. счупване на обекти, изтърване на предмети и др. Във връзка с изричните указания на финансиращия орган, в рамките на проекта не е извършвана дейност по идентификация на личността на който и да е човек. Също така от базата данни е заличена личната информацията, позволяваща идентификацията на конкретен човек, присъстващ в записите.

Изследванията са подкрепени от научноизследователски проект PROMETHEUS (FP7-ICT-214901) по програмата FP7 на Европейската комисия.

[Г7.4]. S.M. Potirakis, N.-A. Tatlas, N. Zafeiropoulos, T. Ganchev, M. Rangoussi (2013). “Assessment of military intercom headsets 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.2012.12.009 (**Thomson Reuters IF2013=1.068**)

В статията се разглеждат проблемите свързани с гласова комуникация в среда с много високи нива на околнния шум, като например в бронетранспортьори, танкове и други бойни превозни средства. В такива условия стандартизираните бойни шлемове с **интерком**, които съдържат вградени слушалки и микрофони, са единственото възможно решение за намаляване на шума при записване и възпроизвеждане на глас. Все така обаче, ефективност и успешност на гласовата комуникация не може да бъде гарантирана безусловно, т.к. надеждността е сравнително ниска.

В статията [Г7.4] се предлага интегриран метод за обективна оценка на електроакустичната пригодност и качество на гласова комуникация, предлагана от специализираните екипировки за гласова комуникация в среда с много високи нива на околния шум. За целта в статията е извършен сравнителен анализ на резултатите от оценяване на интерком екипировки с три различни обективни методи за оценка на качеството на речта, използвайки критерии като точност при автоматично разпознаване на речта (Automatic Speech Recognition -- ASR), както и стандартите за оценка на качеството на речта: Perceptual Evaluation of Speech Quality (PESQ) и индекса за гласова комуникация: Speech Transmission Index (STI). За качествена оценка на метода, който дава най-точни резултати, резултатите от обективния анализ са сравнени със субективни критерии за оценка, като Multiple Stimuli with Hidden Reference and Anchor (MUSHRA).

Предложеният в статията нов метод е използван за оценка на пригодността за гласова комуникация на три комерсиално-достъпни интерком шлема, използвани от военните формирования в Гърция, като същевременно е извършено сравнение с три различни обективни и един субективен метод за оценка. Анализ на експерименталните резултати показва, че методът PESQ осигурява най-подходящия обективен измерител за качество, т.к. е най-добре съгласувана с оценката, получена от субективния метод MUSHRA. Анализът на експерименталните резултати също така показва, че SPL и PESQ измерванията трябва да се оценяват съвместно, т.к. само съвместната им оценка се приближава до субективните оценки.

От извършени субективни полеви тестове, проведени от малък екипаж в условията на нормален режим на бойна машина, един от шлемовете е идентифициран като най-добър с оглед постигане на максимално качество на гласовата комуникация. Същия шлем е предсказан като най-добър и от предложения интегрален метод, което е добра атестация за достоверността на резултатите, получени чрез метода.

[Г7.5]. A. Lazaridis, T. Ganchev, I. Mporas, E. Dermatas, N. Fakotakis, (2012). "Two-stage phone duration modelling with feature construction and feature vector extension for the needs of speech synthesis", *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)

В статията [Г7.5] се издига хипотезата, че чрез автоматично генериране на прибавъчни описатели на речевия сигнал, може да се постигне подобрено автоматично моделиране на продължителността на фонемите в спонтанна реч. В тази връзка е предложен нов метод за моделиране продължителността на фонемите, който намира приложение за подобряване моделирането на прозодиката в автоматичните системи за синтез на реч. Предложеният нов метод умело се възползва от знанието, че различните алгоритми за прогнозиране на продължителността, които работят с общ набор входни описатели, могат да конструират в изхода се нови допълващи се описатели, които, когато се добавят към началния набор, да допринесат за подобряване на общата точност при прогнозиране на продължителността на фонемите.

Предложеният нов метод реализира йерархична двустепенна схема на моделиране. В първия слой се използват множество различни регресивни методи, които служат като автоматични генератори на нови описатели, предсказващи продължителността на фонемите. Във втория слой се извършва същинското предсказване на продължителността на фонемите, като за целта се използва хетерогенен набор от описатели, включващ първоначалния набор от описатели, действащи на входа на първия слой и допълнителните описатели, генерирали в изхода на първия слой. Наборът от описатели, който действа като вход към първия слой, се състои от числови и нечислови лингвистични характеристики, извлечени от текста, а разширеният набор от описатели се получава чрез добавяне на прогнозите за продължителност на фонемите, получени в изхода на генераторите в първия слой на йерархичната схема.

За валидация на издигнатата хипотеза е проведено експериментално сравнение с други методи. Изследвана е работоспособността на метода в експерименталната постановка, включваща осем генератора на описатели в първия слой и десет различни алгоритъма за предсказване на продължителността на фонемите във втория слой. Експерименталният протокол включва отделни експерименти по предсказване дължината на фонемите в спонтанна реч на два езика: американски английски и съвременен гръцки. За целта са проведени експерименти с базата данни KED TIMIT, съдържаща синтезирана реч на американски ,английски и базата данни WCL-1, съдържаща реч на съвременен гръцки език.

Анализът на резултатите от проведените експерименти потвърждават предимството на предложената двустепенна схема на моделиране, която подобрява точността на прогнозите над тази на (i) най-добрия индивидуален предсказател и (ii) двустепенната схема, която само обобщава изходите на предсказателите от първия слой.

Експериментално е установено, че регресията с опорни вектори (SVR) е най-подходящия метод за моделиране на продължителността на фонемите във втория слой на предложената схема. По-конкретно, в сравнение с най-добрия индивидуален предсказател, се наблюдава относителното редуциране на средната абсолютна грешка (MAE) и средната квадратична грешка със съответно 3.9% и 3.9% за KED TIMIT, и 4.8% и 4.6% за базата данни WCL-1.

[Г7.6]. Lazaridis, T. Ganchev, T. Kostoulas, I. Mporas, N. Fakotakis (2010). "Phone Duration Modeling: 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.

Създаването на синтетична реч с високо качество зависи от прецизното моделиране на прозодиката на спонтанната реч, като един от най-важните ключови параметри в случая е продължителността на фонемите. Продължителността на фонемите има важна роля при синтеза на емоционална реч с естествено звучене.

В тази връзка статията [Г7.6] представя критичен обзор на основните методи за моделиране на продължителността на фонемите. Извършен е сравнителен анализ на десет метода за моделиране, използващи: дървета на решенията, линейна регресия, алгоритми за обучение чрез запомняне и др., които през последните десетилетия са били успешно използвани в различни задачи за моделиране. Най-висока точност (най-ниски грешки RMSE и MAE) е постигната при моделиране продължителността на фонемите с дърво на решенията M5p.

Изследвана е възможността за оптимизация на процеса на моделиране чрез прилагане на два метода за подбор на описатели, а именно *RReliefF* и *Correlation-based Feature Selection*. В изследването са включени общо 93 числови и номинални лингвистични описатели, често използвани за характеризиране на текстове, в т.ч. фонетични, фонологични и морфосинтактични, които са докладвани като успешни при моделиране продължителността на фонеми и срички. В статията е изследва практическата полезност на тези методи за моделиране на продължителността на фонемите за база данни от записи на съвременна гръцка емоционална реч, която е представителна за четирите основни категории емоционална реч: гняв, страх, радост, тъга, и неутрална реч.

Анализът на експерименталните резултати показва, че прилагането на процедурата за предварителен подбор на описатели значително подобрява точността на моделиране, независимо от метода за машинно обучение, който е използван за предсказване на продължителността на фонемите. По-конкретно в четири от петте категории емоционална реч, предварителният подбор на описателите допринася за подобряване моделирането на продължителността на фонемите, в сравнение със случая без такъв подбор. Единственото изключение е моделирането на продължителността на фонемите за емоционална реч, изразяваща радост.

[Г7.7]. I. Mporas, **T. Ganchev**, O. Kocsis, N. Fakotakis (2009). "Performance Evaluation of a Speech Interface for 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.

В доклада [Г7.7] е представено изследване на възможностите за успешно речево взаимодействие човек-машина в среда на движещ се мотоциклет. Изследвани са възможностите за успешно автоматично разпознаване на човешка реч, произнесена в условията на физически и когнитивен стрес, характерен при управление на бързо движещ се мотоциклет. Резултатите показват значителна деградация на точността на разпознаване на речта спрямо други видове офис приложения, характеризиращи се с ниски нива на шума.

Изследвани са най-новите и редица традиционни методи за потискане на високите нива на нестационарен шум, (в т.ч. от двигател на мотоциклета, от съприкосновението с пътната настилка, от други превозни средства, околната среда и др.), необходими за създаването на цялостен подход за предварителна обработка на речевите сигнали, специално проектиран за работа в условията на движещ се мотоциклет.

Представено е сравнително изследване на различни методи за подобряване качеството на речевия сигнал при високи нива на нестационарен шум, с които се характеризират акустичните условия при движение с мотоциклет. Тези алгоритми са класирани по отношение на подобрението, което допринася за повишаване на точността при автоматично разпознаване на реч. Експериментите са проведени с помощта на база данни MoveOn и стандартен експериментален протокол. Установено е, че методът с много-лентовото спектрално изваждане допринася за най-висока ефективност при автоматичното разпознаване на реч, в сравнение с други осем метода.

Анализът на експерименталните резултати показва, че изборът на подходяща техника за потискане на шума в речевия сигнал е много съществен за постигане на успешно речево взаимодействие между потребител и системата.

Освен това анализът на експерименталните резултати показва, че няма равностойност между класирането на алгоритмите за шумопотискане въз основа на човешкото възприятие за качеството на речта и въз основа на резултатите, постигнати при автоматично разпознаване на реч.

Изследването е извършено в подкрепа на проекта MoveOn (IST-2005-034753), който е съфинансиран по програмата FP6 на Европейската комисия.

[Г7.8]. 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.

В доклада [Г7.8] е представена общата архитектура на платформата PlayMancer, предназначена за създаване на сериозни компютърни игри, използвани биофеедбек и разширен много-модален потребителски интерфейс. С помощта на PlayMancer могат да се създават компютърни игри, които да се използват като допълнително терапевтично средство при лечение на хронични психични разстройства, в т.ч. поведенчески хранителни разстройства и хронични пристрастявания към хазартни залагания.

Акцентът в доклада е интегрирането на нови функционалности за постигане на много-модалност на потребителския интерфейс и включването на биофеедбек в реално време в управлението на сценариите на играта. В основата на представената разработка са допълване на много-модалната платформа с нови функционалности за разпознаване на жестове, емоционални състояния, състояния на емоционална възбуда и множество жизнени показатели на човешкото тяло, в т.ч. изменениета в сърдечния ритъм, честотата на дишане, повърхностното съпротивление на кожата и др. Това се постига чрез използване на информация, извлечена от реч, видеонаблюдение с камера и физиологични сигнали като ЕКГ, електродермални реакции, дишане и др. информационни канали.

В конкретния случай много-модалността се постига чрез създаването на иновативна платформа, надграждаща архитектурата RavenClaw / Olympus за обработка на спонтанна реч. За целта RavenClaw / Olympus е допълнително разширена с нови модалности и е интегрирана в традиционна платформа за създаване на игри -- UNITY. Платформата PlayMancer разчита на модулна архитектура, където компонентите за обработка на потоците от данни от индивида от отделните модалности и източниците на входни въздействия (мишка, клавиатура, джойстик) си взаимодействват чрез централен хъб в ядрото на Olympus. Това позволява синхронно или алтернативно използване на входно / изходните модалности. По този начин в допълнение към традиционните входове / изходи, използвани в игрите (джойстик, клавиатура, мишка, дисплей), в архитектурата на PlayMancer платформата са интегрирани също възможности за обработка и извличане на информация от реч, допир, сензори за регистриране на физиологични сигнали и сензори за проследяване на движение.

Тези допълнителни възможности позволяват използването на нови подходи при създаването на компютърни игри, включително и с осигуряване на биофеедбек в реално време. Въвеждането на биофеедбек при разработката на компютърни игри позволява разработването на игри с подобрени преживявания за средностатистически геймър, но по-важното е, че те осигуряват средства за развитие на нов тип сериозни игри за борба с различни хранителни разстройства и пристрастявания или за целите на физическата рехабилитация след сериозни травми и заболявания.

Изследванията, представени в доклада, са финансиирани по проект PlayMancer (FP7-ICT-215839), който е частично финансиран от Европейската комисия в рамките на Седма рамкова програма.

**[Г7.9]. 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.**

В статията [Г7.9] е представен нов метод за подобряване прецизността при сегментация на речеви сигнали до фонеми. Предлаганият метод се основава на добре известния алгоритъм на Витерби за времево подравняване на сигналите и използва границите между фонемите, които са определени чрез множество предсказатели, използващи различни методи за параметризиране на реч. По-конкретно, най-добрата предсказана позиция на фонемата, за всеки тип граница се използва като началната точка за започване на времевото изравняване с алгоритъма на Витерби, с цел по-прецизно предсказване на следващата граница.

Прецизността на предложението метод е оценена с помощта на стандартен експериментален протокол, използващ базата данни от записи на американска реч TIMIT. За моделиране на фонемите са използвани скрити модели на Марков (HMM ToolKit -- HTK) с б състояния и еднопосочна организация от ляво надясно. За всяко от шестте състояния са изследвани модели, създадени със смес от Гаусови функции, използващи 2, 3, 4 или 6 компонента.

За обучението на отделните предсказатели на границите на фонемите са използвани общо 7 различни метода за извлечение на описатели от речевия сигнал. От тях 4 използват преобразуване на Фурье, като първите три са добре познати (LFCC, MFCC, PLP CC), един е по-нов и малко проучен метод (HFCC-E), който досега не е изследван в приложения за сегментиране на реч. Други 3 метода, които използват дискретно преобразуване с уейвлет пакети (DWPT) са нови (MWP-ACE, WPF, SBC) и досега не са изследвани в приложения за сегментиране на реч.

Сравнителен анализ на работата на седемте метода показва, че най-висока прецизност на предсказване на границите на фонемите се получава при използване на описатели HFCC-E, които превъзхождат традиционно използванието MFCC.

Използвайки същия експериментален протокол е изследвана прецизността на предложението нов метод. Анализът на експерименталните резултати показва предимството на предложението метод за сегментиране на речта спрямо най-широко използвания базов метод (MFCC) с до 3.5%, за всички допустими толеранси на границите на фонемите.

В сравнение с най-добра индивидуален предсказател, използващ HFCC-E описателите, най-голямо подобреие на прецизността е отчетено за толеранс от 20 милисекунди, при който се наблюдава подобряване на абсолютната точност на сегментиране с 0.7%.

**[Г7.10]. I. Mporas, T. Ganchev, O. Kocsis, N. Fakotakis (2010). "Speech Enhancement for Robust Speech Recognition in Motorcycle Environment", International Journal on 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)**

В статията [Г7.10] са представени задълбочени изследвания по усъвършенстване на метод за автоматично разпознаване на реч, произнесена в условията на значителен физически и когнитивен стрес. Този метод е част от речеви интерфейс, който е ключов компонент от система за информационно обслужване на мотоциклисти полицаи. Изследванията, представени в статията, надграждат и доразвиват метода за шумопотискане, представен в доклада [Г7.7]. Новостите са две: освен шумопотискане, в настоящата статия [Г7.10] се въвеждат: (i) процедура за адаптация на акустичния модел на системата за разпознаване на реч към акустичните условия, характерни за мотоциклет в движение и (ii) рестриктивна граматика, която е адаптирана към нуждите на конкретното потребителско приложение.

Проведена е експериментална оценка на системата за автоматично разпознаване на реч: (i) с използване и без използване на шумопотискане, (ii) с адаптация и без адаптация на акустичния модел спрямо съответната работна среда, (iii) с използването на *bigram* и *trigram* лингвистичен модел. Оценена е работата на осем метода за шумопотискане в комбинация с и без адаптация на акустичния модел на системата за автоматично разпознаване на реч. Проведени са изследвания за реч записана в офис среда с ниски нива на шума и в среда на уличен трафик с високи нива на шума, и при движещ се мотоциклет.

Анализ на експерименталните резултати показва, че когато не се използва шумопотискане и адаптация на акустичния модел, разпознаването на реч в среда на движещ се мотоциклет е с абсолютна точност от 24%, което означава практически неизползваема система за разпознаване на реч. След прилагане на метод за шумопотискане точността на разпознаване се повишава значително, като най-добрия резултат се получава за шумопотискане с метод, базиращ се на много-лентовото спектрално изваждане (M-BAND), при който точността на разпознаване на речта достига до около 55%, но и това е крайно недостатъчно за използване в практически приложения.

При използване на акустични модели, които са адаптирани към работната среда, най-добра точност при разпознаване на реч в среда на движещ се мотоциклет (над 86.6%) се постига чрез метода за шумопотискане, използващ Бейсов оценител на амплитудата на речевия спектър (STSA-WCOSH). Други три метода за шумопотискане също показват много добри резултати при използване на адаптиран акустичен модел, постигайки точност на разпознаване от почти 85%. В сравнение с резултатите, представени в [Г7.7], в текущата статия се докладва подобреие с над 30% на точността на разпознаване на реч (по критерии Word Recognition Rate -- WRR).

Анализът на експерименталните резултати показва, че за конкретното приложение е от изключително голяма важност да се използва правилната комбинация от (i) метод за шумопотискане, (ii) акустичен модел, който е адаптиран към работната среда и (iii) лингвистичен модел, целево разработен за конкретното приложение.

Изследването е извършено в подкрепа на проекта MoveOn (IST-2005-034753), който е съфинансиран по програмата FP6 на Европейската комисия.

[Г7.11]. 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, Czech Republic, pp. 4732-4735.

В доклада [Г7.11] е представен нов метод за определяне продължителността на фонемите, посредством обединяване на предсказанията на група от разнородни модели чрез регресия с опорни вектори (SVR). В предложения метод се използва двустъпална архитектура на предсказване, като в първия слой се създават множество индивидуални модели за предсказване продължителността на фонемите, които използват еднакъв набор от описатели на речта. Всеки от осемте модела в първия слой е създаден като се използва различен метод за машинно обучение, в т.ч. линейна регресия, регресивни дървета, адитивна регресия и др. Предложеният метод разчита на факта, че различните методи за машинно обучение използват различни допускания при моделирането на разпределението на описателите на речта, при което се очаква да грешат по различен начин. В тази връзка методът за машинно обучение, използван за обединяване на предсказанията, получени от отделните модели, може да бъде обучен да коригира грешките на индивидуалните модели. По такъв начин допускаме, че може да се подобри прецизността даже и на най-прецизния индивидуален предсказател, независимо по какъв метод е създаден.

Сравнителното изследване на прецизността на индивидуалните предсказатели и на предложения двустъпален метод е извършено в общ експериментален протокол, използваш базата данни KED TIMIT със записи на американска реч и базата данни WCL-1, съдържаща записи реч на съвременен гръцки език.

Анализът на експерименталните резултати показва, че предложеният нов метод допринася за подобряване точността на най-добрия индивидуален модел. Постигнато е относително намаляване на средната абсолютна грешка (MAE) и средна квадратична грешка (RMSE), с 1.9% и 2.0% за KED TIMIT, и съответно 2.6% и 1.8% за WCL-1.

За оценка на въздействието на наблюдаваното подобреие на прецизността при определяне на границите на фонемите, върху възприятието на синтетичната реч е проведен субективен MOS тест.

Анализът на резултатите от MOS теста показва, че подобрението на точността, постигнато чрез предложения метод за моделиране на продължителността на фонемите, допринася за подобряване на естествеността на синтетичната реч спрямо най-прецизния индивидуален метод и спрямо втория най-добър индивидуален метод.

[Г7.12]. 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.

В доклада [Г7.12] се представя дизайна и изпълнението на базата данни от реч Vergina, която е разработена в подкрепа на научноизследователската и развойна дейност, свързана със създаването на статистически параметрични системи за синтез на реч за съвременен гръцки език.

При проектирането на фонетичния и лингвистичния баланс на Vergina е използван статистически анализ на текстов корпус на около 5 miliona думи, събрани от вестникарски статии, периодични издания и параграфи от литературата, които са обработени статистически с цел избор на изречения и фрази, които са балансирани фонетично и лингвистично. Широкият обхват и фонетично съдържание на записите в базата данни прави Vergina подходяща за различни области на приложение. Описан е експерименталният протокол, използван при събирането на базата данни и процедурата за анотация.

Докладът представя структурно описание на базата данни, включващ статистическо описание, резултати за най-често срещаните думи, най-често срещаните двойки фонеми и др. Накратко базата данни Vergina се състои от около 4 часа реч, записана в професионално аудио студио. Vergina се състои от приблизително 3,000 фонетично балансираны фрази и изречения на съвременен гръцки език.

Транскрипцията на речта в базата данни и определянето на границите между фонемите е извършено с използване на автоматичен метод за сегментиране на реч, използваш скрити модели на Марков (HMM) с хибридно обучение, след което са въведени ръчни корекции за подобряване прецизността на анотациите.

[Г7.13]. 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.

Разработването на системи за синтез на реч (*Text-to-Speech*), които са създадени въз основа на база данни от записи на реч е тясно свързано с наличието на фонетични транскрипции и прецизно определени граници между фонемите. Тъй като ръчната сегментация на речта е трудоемка и до известна степен субективна, възниква необходимостта от разработване на автоматични методи за фонетично сегментиране на речта. Наличието на автоматични методи би подпомогнало за значително съкращаване на времето за разработка и усъвършенстване на системите за синтез на реч.

В настоящия доклад [Г7.13] се изследва полезнотта на два различни метода за автоматична сегментация на реч, с цел автоматично определяне на границите на фонемите за нуждите на създаването на система за синтез на реч на съвременен гръцки език.

Първият метод използва традиционен подход за сегментация с помощта на система за автоматично разпознаване на фонеми, използваща традиционните методи за моделиране със скрити модели на Марков (Hidden Markov Models, HMM).

Вторият метод е модификация на метода, предложен в доклада [Г7.18], при който се използва хибридна архитектура с вградено и изолирано обучение на HMM модели за автоматично сегментиране на реч в контекста на емоционална реч от един диктор. Новите аспекти, които изискват променена конфигурация на метода спрямо предложения в [Г7.18], се състоят в използването на база данни със записи на съвременен гръцки език от един диктор, произнасящ фрази и изречения с емоционална окраска, изразяваща пет различни емоционални състояния.

Анализът на експерименталните резултати, сравняващи прецизността на сегментиране показва, че методът използващ хибридно обучение на HMM, предлага значително подобреие в точността на сегментирането на речта, в сравнение с базовия HMM метод.

Анализът на резултатите от субективния MOS тест за оценка на качеството на синтезираната реч показва, че хибридният метод води до значително по-добро качество на синтетичната реч в сравнение с базовия метод, като качеството на синтетичната реч е близко до това, получено при използване на ръчно създадени фонетични транскрипции и определяне на границите.

Основно предимство при използване на метода с хибридно обучение е, че води до намаляване на времето за разработка на *Text-to-Speech* системите при запазване на удовлетворително качество на синтезираната реч.

Изследването е извършено в подкрепа на проекта MoveOn (IST-2005-034753), който е съфинансиран по програмата FP6 на Европейската комисия.

**[Г7.14]. 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.**

В доклада [Г7.14] са представени резултатите от изследване на работата на система за автоматично разпознаване на реч на съвременен гръцки език. Системата за автоматично разпознаване на реч е основна част от диалогова система за гласово управление на домашни уреди в среда на интелигентен дом. За разлика от гласовите системи, където е необходимо потребителят да произнася определен брой зададени команди за контрол и управление, в настоящата система потребителите могат да използват спонтанна разговорна реч.

В настоящото изследване, системата за разпознаване на реч е разработена с помощта на софтуер с отворен код, HMM ToolKit (HTK), който работи на всеки домашен компютър. С помощта на HTK и базата данни SpeechDat(II)-FDB-5000-Greek е създаден акустичен модел, който използва HMM моделиране с 5 състояния по топологията на Бакис. Граматиката съдържа 73 думи – команди и техните синоними, най-често използвани думи, за гласово управление на телевизор, телефон, DVD плейър, пералня др. и включва и общ модел (*garbage model*), който да е представителен за спонтанна реч, която не е от интерес за управлението на домашните уреди.

Работата на системата за автоматично разпознаване реч е изследвана за 10 потребители при различни шумови характеристики на околната среда, за два различни вида микрофони – обикновен PC микрофон и висококачествен безжичен микрофон AKG C444, който е снабден с метод за автоматично потискане на шума. По време на експерименталната валидация на системата в условия на различни отношения сигнал-шум (SNR 6dB, 12 dB, 30 dB) и висококачествена реч, при използването на микрофон AKG C444, прецизността на разпознаване на командите е в обхвата 94-96% и 98%, съответно. Въпреки различната погрешност при разпознаване на думите и с двата микрофона е отчетено 100% изпълнение на задачите във всички сценарии.

Използването на системата за разпознаване на реч с отворен код в комбинация с използването на евтин микрофон за близко говорене, позволява евтино решение за домашна автоматизация. Потребителят получава обратна връзка, като просто наблюдава оперативното състояние на съответния уред.

**[Г7.15]. I. Mporas, O. Kocsis, T. Ganchev, N. Fakotakis (2009). "A Collaborative Speech Enhancement Approach For Speech Recognition In Motorcycle Environment", Proc. of the 16th International Conference on Digital Signal Processing, DSP 2009, Santorini, Greece. 5-7 July, 2009. Article number 167.**

В доклада [Г7.15] се предлага нов колаборативен метод за шумопотискане, с цел оптимизиране ефективността на работа на система за автоматично разпознаване на реч, която е важна част от много-модален диалогов интерфейс за информационна подкрепа на специализирано подразделение полицията, състоящо се от полицаи мотоциклисти.

Предложеният нов метод използва множество независими канали за разпознаване на реч, всеки от които работи с различен метод за шумопотискане, които в последствие се обединяват с помощта на метод за машинно обучение.

Извършено е изследване на различни конфигурации на системата, включително методите за предварителна обработка на речевия сигнал и методите за моделиране и разпознаване на реч при високи нива на шум и високи нива на физически и когнитивен стрес.

Извършен е сравнителен експериментален анализ на практическата полезност на шест различни методи за шумопотискане и резултатите са сравнени с тези на предложения нов метод. Експериментално е установено, че независимо от използвания метод за обединение на отделните канали, използването на колаборативния метод води до значително подобряние на точността на разпознаване на речта.

Анализ на експерименталните резултати показва, че използването на предложения колаборативен метод за разпознаване на реч при високи нива на шума, води до подобряване на точността по отношение на процента на разпознати думи с 8% спрямо най-добрия от индивидуалните канали за разпознаване на реч.

Методите Adaboost.M1(J48) и IBk позволяват най-добри резултати при обединението на индивидуалните канали за разпознаване на реч, при което точността на разпознаване достига до около 95%. Такава точност на разпознаване се счита за приемлива в конкретната работна среда на движещ се мотоциклет, високи нива на шум и високи нива на физически и когнитивен стрес.

Изследването е извършено в подкрепа на проекта MoveOn (IST-2005-034753), който е съфинансиран по програмата FP6 на Европейската комисия.

[Г7.16]. R. Saeidi, H.S. Mohammadi, **T. Ganchev**, R.D. 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.

В доклада [Г7.16] се изследва прецизността на система за верификация на диктори, независимо от лингвистичното съдържание на речта. Системата използва метод за моделиране на разпределението на описателите на речта със сортирана смес от Гаусови функции (SGMM), който позволява по-бързо изчисляване на резултата спрямо традиционните модели, използвани със смеси от Гаусови функции (GMM). Цел на изследването е да се оцени възможността за подобряване на точността на SGMM, чрез използване на методи за нормализация на описателите като Cepstral Mean Subtraction (CMS) и Dynamic Range Normalization (DRN). Макар ефективността на тези нормализации да е успешно използвана в системите за разпознаване диктори с UBM-GMM, досега тяхната ефективност не е била изследвана в контекста на SGMM.

Оценката на различни конфигурации на системата за верификация на диктори е извършена с помощта на стандартен експериментален протокол, дефиниран от NIST през 2002 година специално за годишната кампания за разпознаване на диктори. За целта е използвана базата данни реч NIST 2002, съдържаща реч от 330 диктора. За създаване на универсалния модел UBM със смес от 1024 Гаусови компонента са използвани 200 диктора от базата данни реч NIST 2000.

Анализ на експерименталните резултати показва, че нормализацията и с CMS и с DRN подобрява точността на базовата система, използваща UBM-GMM моделиране с 1.1% и снижава загубата на точност в системата с ускорено изчисляване на резултата чрез SGMM с 1.3% до 4.2%, при ускоряване на изчисленията между 2 и 79 пъти. Комбинираното използване на двата вида нормализация CMS+DRN допълнително подобрява точността с още 0.3% до 1.3%.

[Г7.17]. R. Saeidi, H.R.S. Mohammadi, T. Ganchev, R.D. Rodman (2008). "Hierarchical mixture clustering and its application to GMM based text independent speaker identification", *Proc. of the International Symposium on Telecommunications*, IST2008, Teheran, Iran, Aug. 27-28, 2008, pp.770-773.

В доклада [Г7.17] се разглеждат възможностите за подобряване на мащабируемостта при създаване на системи за идентификация на диктори с голям брой клиенти. В доклада се предлага нов метод за йерархична кълстеризация на моделите на отделните диктори (GMM-HMC), като се взима в предвид близостта им до универсалния акустичен модел (UBM). Експериментално е показано, че с помощта на предложения метод може да се постигне намаляване на общия брой и сложността на изчисленията и ускорено получаване на крайния резултат.

Анализът на резултатите от сравнителни експерименти показват, че с помощта на GMM-HMC може да се постигне по-точно групиране на дикторите отколкото с метода, използваш сортиран GMM със същата степен на ускорение. Системата е тествана за основния GMM-HMC и три модифицирани варианта. Във всички случаи методите за намаляване на броя на изчисленията понижават точността на идентификация на дикторите с 3% - 5% спрямо базова система при понижаване броя на изчисленията около 3-4 пъти. Анализът на резултатите показва загуба на точност с около 3.3% при намаляване на броя на изчисленията 3.2 пъти.

Въз основа на експерименталните резултати е направено заключението, че по-нататъшното развитие на метода за йерархична кълстеризация (GMM-HMC), има потенциал да предложи възможности за създаване на системи за идентификация на диктори с по-високи точност и по-ниска изчислителни загуби.

- [Г7.18]. 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.

В доклада [Г7.18] се предлага хибридна архитектура за автоматично подравняване на речевите сигнали към съответстващите им поредица от фонеми, без да се използва предварителна информация за границите на фонемите. Това осигурява голямо предимство пред предишни методи, при които се изисква първоначален набор (*bootstrap dataset*) от ръчно изготвени фонемни транскрипции, при които ръчно са поставени маркери за границите на фонемите. Този първоначален набор от данни е необходим, за да се извърши моделиране на фонемите и първоначално определяне на границите им. Ръчното определяне на границите на фонемите е трудоемък процес, при който има известна доза субективизъм, поради което се налага един и същи текст да бъде транскрибиран от няколко лингвиста. Чрез предложения нов автоматичен метод, нуждата от ръчно създадени анотации се елиминира, при което отпада несигурността от човешката субективност. Предложеният метод е полезен за приложения, при които се изисква автоматично създаване на фонетични транскрипции на текстове, в това число и в системите за автоматичен синтез на човешка реч.

Предложеният нов подход е илюстриран чрез нов автоматичен метод за сегментация на реч, който комбинира ефективността на вградените методи за обучение на фонемни модели, реализирани чрез скрити модели на Марков (Hidden Markov Models, HMM) и високата прецизност на изолираното обучение на HMM. Тъй като в предложения нов метод не се използват ръчно създадени фонетични транскрипции, всички HMM модели се инициализират равномерно, с еднакви стойности на параметрите. След това параметрите на всички HMM модели се преизчисляват и донастройват едновременно с помощта на Baum-Welch алгоритъма. Това се повтаря определен брой пъти до достигане на устойчивост на параметрите на моделите.

С помощта на стандартен експериментален протокол, използващ базата данни от американска реч, TIMIT, която съдържа записи на реч от 640 человека, е изследвана точността на сегментация. В среда на множество различни диктори е извършено сравнително изследване на прецизността на предложения нов метод спрямо тази на широко използван базов метод. Изследвани са различни конфигурации на параметрите на предложения нов метод, в това число при моделиране с HMM с 5 и с 6 състояния за всяка фонема, които са най-широко използвани модели в системите за сегментиране или разпознаване на реч. Направено е сравнение с традиционен базов метод при същия брой състояния и при използване на вградено и изолирано обучение на моделите. Анализът на експерименталните резултати показва, че след малък брой итерации (обикновено между 15 и 20), прецизността на сегментация на речевия сигнал нараства с над 6%, независимо дали се използват HMM модели с 5 или с 6 състояния.

Изследванията, представени в доклада, са частично подкрепени от проекта LOGOS (ЕНГ-102), който се финансира от Генералния секретариат за Научни изследвания и технологии на Гърцкото министерство на развитието.

[Г7.19]. S. Ntalampiras, I. Potamitis, **T. Ganchev**, 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*, LREC-2008, Morocco, May 28-30, 2008. pp.1288-1291.

В доклада е представен концептуалния проект и реализацията на база данни от аудио записи, създадена в подкрепа на изследванията по автоматично акустично наблюдение на публични пространства. База данни е създадена като част от по-мащабен проект, насочен към създаване на методи и средства за автоматично откриване и интерпретация на необичайни събития, които водят до застрашаващи живота ситуации или имуществени щети.

Аудио записите съдържат гласови събития и звуци на околната среда, които обикновено се срещат по време на нетипични ситуации. Основното предназначение на този набор от данни е да предоставят ресурсите, необходими за създаване на акустични модели и за обучение на автоматични методи за разпознаване и интерпретация на аудио събития, с цел автоматично откриване на опасни ситуации. Тези функционалности спомагат за повишаване на сигурността в обществената среда, наблюдането на която в момента изисква денонощен непрекъснат човешки надзор.

Базата данни от аудио записи се състои от три части:

- Фаза I, представлява компилация от различни професионални колекции от звукови ефекти,
- Фаза II съдържа записи, получени от екшън и драматични филми, и
- Фаза III съдържа реални гласови реакции, свързани с истински извънредни събития, които са извлечени от телевизионни предавания, новини на живо, по време на радиопредавания, документални филми и т.н.

В доклада са представени методиката за анотация на аудио записи и статистически анализ на набора от данни, както и резултати от класификация на 10 категории звукови събития от Фаза I, в т.ч. звук от изстрел с огнестрелни оръжия, счупване на стъкла, изтърване на предмети и др. ситуации, типични за опасни и незаконосъобразни действия.

При анализа на представените резултати, които са счетени за предварителни е отчетена необходимостта за допълнително задълбочено проучване, което да спомогне за разбиране за различията между нормални и нетипични събития от акустична гледна точка, с цел правилна интерпретация на ситуацията.

Изследванията са частично подпомогнати от проект Prometheus-214901 "Prediction and Interpretation of human behaviour based on probabilistic models and heterogeneous sensors", финансиран по Седма рамкова програма (FP7) на Европейската комисия.

[Г7.20]. 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.

В доклада [Г7.20] се предлага нов метод за извличане на описатели на речта, чрез който се изчислява нов тип описатели, Overlapping Discrete Wavelet Packet Features -- ODWPFs, наречени *описатели с припокриващи се уейвлет пакети*, оптимизирани за верификация на диктори. Тези описатели са целево проектирани за по-успешно разграничаване на човешки гласове, което се постига чрез специално създадена методика за оценяване и подбор на уейвлет дървета за разлагане на сигнала, осигуряващи предимство при верификацията на диктори. Методиката за подбор на уейвлет дървото на разлагане използва обективен критерий – точността при верификация на диктори – за автоматично оценяване на голям набор от ортогонални или неортогонални базисни функции. В резултат от прилагането на методиката се избира най-успешното уейвлет дърво на разлагане, което означава най-добър набор от базисни функции, който да подчертава различията между човешките гласове. В този случай ортогоналното преобразуване с уейвлет пакети се явява само един частен случай -- при липса на припокриване между избраните базисни функции.

Използвайки (i) новосъздадената методика за откриване на оптимално уейвлет дърво на разлагане и оптимален набор от базисни функции и (ii) новия метод за изчисляване на описатели са изчислени нов вид описатели на речта, WP-2011, които използват неортогонално разлагане с уейвлет пакети. За нуждите на сравнителния анализ е създадено подобно уейвлет дърво на разлагане без препокриване между честотните ленти и са изчислени нов набор от описатели наречени WP-0000.

За изследване на пригодността на новия набор от описатели (WP-2011) е следван стандартен експериментален протокол, дефиниран от организаторите на състезанието 2001 NIST SRE, което е насочено в подкрепа на развитието на технологиите по разпознаване на диктори. За целта е използвана част от базата данни от речеви сигнали 2001 NIST SRE, която е записани през мобилната телефонна мрежа на САЩ и дефинираните от организаторите правила за използването и. Следвайки стандартния експериментален протокол *one-speaker detection*, дефиниран от NIST е извършен сравнителен анализ между предложените описатели, WP-2011 и други пет вида описатели, от които два наскоро създадени (WP2-Sarikaya и WP3-Farooq-Data), използващи преобразуване с уейвлет пакети, два традиционни (MFCC FB-20 (HTK) и MFCC FB-32), използващи време-честотен анализ с кратковременно преобразуване на Фурье, както и WP-0000, при който се изчисляват уейвлет описатели, като се използва ортогонален набор от базисни функции.

Анализът на експерименталните резултати с предложения новия тип описатели на речта (WP-2011), оптимизиран за нуждите на верификация на диктори, показва относително намаляване на погрешността при разпознаване с 27%, 22% и 15% спрямо описателите WP3-Farooq-Data, MFCC FB-32, и WP2-Sarikaya. Използването на неортогонално преобразуване с уейвлет пакети допринася за относително понижаване на погрешността при верификация на диктори с 3% спрямо еквивалентното ортогонално уейвлет преобразуване.

**[Г7.21].** 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, September 2004, pp.137-145.

В доклада [Г7.21] е представен нов стохастичен модел за анотация на частите на речта в текстове на естествен език, при който се използва нов хибриден класификатор В-PNN, комбиниращ Бейсова мрежа и Вероятностна невронна мрежа.

Докато предишните статистически подходи, като например скритите модели на Марков (HMM), се основават на теоретични предположения и допускания, които не винаги са в сила за естествения език, в настоящия доклад се предлага методология, която комбинира фундаментални елементи от два метода за машинно обучение с цел осигуряване на контекстуална информация при класификацията. За целта е създадена Вероятностна невронна мрежа (PNN), използваща допълнителна контекстна информация чрез Бейсово представяне на знания, за да се обучи устойчив класификатор, който да разграничава успешно частите на речта. Обучението се извършва с малък лексикон, който съдържа думите, които принадлежат към несклоняеми части на речта. Тази информация се допълва от статистически параметри, генериирани от обучението на Бейсови мрежи и резултата се подава към Вероятностната невронна мрежа за осигуряване на контекстуална информация.

Сравнителната оценка на предложения хибриден класификатор, В-PNN, с класификатор реализиран с Бейсова мрежа (BN) или със скрити модели на Марков (HMM) е извършена в помощта на общ експериментален протокол. За целта са използвани базите данни с текстове на съвременен гръцки език ILSP Greek Corpus с 250 000 думи и ESPRIT(291-860) Greek Corpus с 120 000 думи. Всички експерименти са извършени с десетократна валидация (10-fold cross validation), т.е. десетократен избор на обучаващи и тестващи набори от данни.

Анализът на експерименталните резултати показва, че предложенияят хибриден класификатор В-PNN има по-ниска погрешност при класификация на частите на речта, средно с 2.5% - 3.5% спрямо Бейсова мрежа (BN) и скритите модели на Марков (HMM). Като цяло предложенияят хибриден класификатор В-PNN показва най-добри резултати и на двете бази данни, като грешката е под 4%. Такава прецизност се счита задоволителна за анотация на частите на речта.

**[Г7.22]. M. Siafarikas, T. Ganchev, N. Fakotakis (2004). "Objective Wavelet Packet Features for Speaker Verification", Proc. of the 8th International Conference on Spoken Language Processing, InterSpeech-2004 - ICSLP, Oct 4- 8, Jeju, Korea, 2004, pp. 2365-2368.**

В доклада [Г7.22] е представен нов метод за изчисляване на описатели на речеви сигнали, които са оптимизирани за нуждите на автоматично разпознаване на диктори. Предложеният метод използва преобразуване с уейвлет пакети изчисляване на кратковременната енергия в много голям брой честотни ленти (128 или 256). След това се извършва систематична оценка на приноса на отделните честотни ленти по отношение на важността при разпознаване на диктори. Систематичният подбор позволява честотните ленти, които са отчетени като по-значими при разпознаването на диктори, да бъдат моделирани с по-добра разделителна способност по честота. По този начин се създава уейвлет дърво за разлагане на речевия сигнал, което е специално проектирано за по-добро разграничаване на гласовете на дикторите.

Експерименталният протокол използва базата данни със записи на телефонна реч (2001 NIST SRE), която е целево създадена за оценка на качеството на различни методи за разпознаване на диктори. В експериментите е използван набора записи, направен през клетъчните мобилни мрежи на САЩ. Всички експерименти са извършени с помощта на система за верификация на диктори, която използва Вероятностни невронни мрежи, при която се създава отделен детектор за всеки диктор.

Извършен е сравнителен анализ и оценка на практическото значение на предложениия метод с три други вида съвременни описатели на речта, базирани на уейвлет пакети и традиционните описатели на речта, познати като кепстрални коефициенти, изчислени чрез банка филтри с 32 филтъра с Мел разпределение на централните честоти на филтрите (MFCC).

Анализът на експерименталните резултати показва, че предложеният набор от обективно оптимизирани уейвлет описатели превъзхожда MFCC FB-32 с около 2.4%. Това предимство се дължи на оптимизирания дизайн на уейвлет дървото на разлагане, което е определено по обективен начин, специално за нуждите на разпознаването на диктори.

Изследванията публикувани в доклада са частично подкрепени от проект INSPIRE (IST-2001-32746) "Infotainment management with Speech Interaction via Remote microphones and telephone interfaces" съфинансиран по програмата FP5 на ЕС.

### 7.3 Резюмета за публикациите от група „Г.8“

[Г8.1]. Н. Дуков, Т. Ганчев (2018). “Нови невронни модели на ЛРВНМ, базирани на RELU активиращи функции”, списание *Компютърни Науки и Технологии*, Година XVI, Брой 1/2018, ТУ-Варна, стр.135-142.

В статията [Г8.1] се предлага нова разновидност на локално рекурентната вероятностна невронна мрежа (LRPNN), при която невроните в локално рекурентния слой се реализират с нова активираща функция от семейството на ReLU, наречена SatReLU. Активиращата функция SatReLU е създадена с цел, да се комбинират положителните характеристики на добре известните Leaky ReLU и Clipped ReLU, които вече са били изследвани в доклада [Г8.7]. С тази модификация на невроните от локално рекурентния слой се елиминира основен недостатък при апаратна реализация на LRPNN, понеже традиционно използваната сигмоидална активираща функция е нелинейна, със значителната изчислителна сложност, поради което е трудна за имплементиране в FPGA чип.

Предимствата, произтичащи от използването на предложените нови модели на локално рекурентни неврони, са изследвани с помощта на експериментална постановка за разпознаване на негативни емоционални състояния от физиологични сигнали, реализирана с помощта на базата данни DEAP. В експериментите по разпознаване на негативни емоционални състояния са използвани 32-канални ЕЕГ записи на 10 человека от базата данни. От ЕЕГ сигналите са изчислени описатели, които представлят нормализираната кратковременна енергия за малки припокриващи се отрезъци от сигнала. Описателите, изчислени в даден момент от времето за всички канали, се подават на входа на невронната мрежа, като вектор състоящ се от 32 стойности.

Извършено е сравнително експериментално изследване на работоспособността на модифицираната локално рекурентната, вероятностна, невронна мрежа (LRPNN) при различни модели на невроните в локално рекурентния слой. Изследвани са модели на неврони с: (i) сигмоидална, (ii) ReLU, (iii) SatReLU, (iv) Leaky ReLU, (v) Clipped ReLU активираща функция.

Анализ на експерименталните резултати показва, че използването на предложената нова активираща функция за невроните от локално рекурентния слой на LRPNN носи преимущества, изразявачи се в подобрена точност при класификация, намалена сложност на изчисленията и подобрена скорост на обучение на рекурентния слой. Подобрена точност при класификация е постигната за Leaky ReLU с  $\text{const} = 0.07$  и предложената SatReLU с  $\text{const} = 80$ .

Активиращата функция SatReLU демонстрира по-добри показатели от гледна точка на предвидимостта спрямо останалите изследвани функции, което с оглед на малката разлика ( $\approx 1\%$ ) с Leaky ReLU в постигнатата точност я позиционира като нейна добра алтернатива.

Друго основно преимущество при използването на ReLU и нейните разновидности е скоростта на сходимост при обучението на невронната мрежа, която е приблизително 3 пъти по-висока отколкото при използването на сигмоидалната функция. Тези преимущества показват, че предложените модификации на LRPNN са особено подходящи при апаратна реализация с помощта на цифрови платформи от типа на FPGA.

[Г8.2]. 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.

В статията [Г8.2] е изследвана приложимостта на 10 статистически описатели на ЕЕГ сигналите за откриване на негативни емоционални състояния. Тези описатели се изчисляват от ЕЕГ сигнала директно във времевата област: *средна стойност, стандартно отклонение, кратковременна енергия, модул на средната стойност на първата разлика на сигнала, модул на средната стойност на втората разлика на нормализирания сигнал, модул на средната стойност на втората разлика на нормализирания сигнал, активност, мобилност и комплексност*. При изчисляването на средна стойност на модула на първата и втората разлика на отчетите на нормализирания сигнал, ЕЕГ записите се подлагат на статистическа нормализация. След като вече са изчислени, десетте описатели се подлагат на статистическа нормализация до нулева средна стойност и единично стандартно отклонение.

Изследването обхваща 4 сценария за разделяне на данните на базата на критериите за оценка на стимулите: „възбуда“, „доминантност“, „валентност“ и „харесване“. При това се формират 4 набора от данни, които позволяват да се откриват различни аспекти на емоционалност за едни и същи участници, чрез едни и същи записи на ЕЕГ сигнали. Първият набор позволява да се следи нивото на емоционална възбуда, вторият - доколко чувството е завладяващо (или степента, в която може да бъде контролирано), третият - дали чувството е негативно или друго (позитивно, неутрално) и четвъртият - доколко чувството е свързано с харесване и симпатия.

За провеждането на експерименталната оценка на описателите е използвана автоматизирана система за детекция на негативни емоционални състояния, която използва класификатор с опорни вектори (SVM). Използвани са данните на 10 участника от базата данни DEAP. Във всеки сценарий за всеки участник е обучен отделен детектор. В зависимост от предоставените анотации, данните на участниците се разделят на две групи – „целеви клас“ и „други“. В случай на „валентност“ това са „негативни емоции“ и „други“.

Получените резултати показват, че описателите средна стойност на модула на първата и втората разлика на отчетите на нормализирания сигнал, показват най-добри класификационни резултати – 68.7% и 68.6%. Най-ниска точност е получена при използването на описател средна стойност на сигнала – 62.9%. От анализа на средната точност може да се види, че по-комплексните описатели демонстрират по-високи резултати. Допълнително може да се посочи, че кратковременна енергия на сигнала е единственият описател, който освен средна стойност на модула на първата и втората разлика на отчетите на нормализирания сигнал е позволил успешно създаване на устойчиви модели за 80% от случаите.

Въпреки зависимостите, наблюдавани при анализа на средната точност, предвиждането на поведението на описателите за отделните участници остава трудно. Причината за наблюдаваното непостоянство в полезнотта на описателите се дължи предимно на спецификите на ЕЕГ сигналите за отделните индивиди и на нееднаквото качество на ЕЕГ записите.

[Г8.3]. 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.

Успешната аппаратна реализация на Вероятностна невронна мрежа (PNN) зависи както от оптимизация на трудните за изпълнение нелинейни функции в невроните, така и от постигане на устойчив и реализуем процес на обучение. Представеното в доклада изследване е насочено към намиране на подходящ метод за обучение на PNN, който да е удобен за аппаратно изпълнение с FPGA и по възможност - с икономична реализация в чипа.

В тази връзка, в доклада [Г8.3] се изследват три еволюционни метода за обучение на Вероятностната невронна мрежа, с цел намиране на приемлив компромис между точността на класификация, изчислителната сложност, необходимия обем памет и размера на популацията от елементарни агенти, използвани при оптимизацията. Подбраните три типа оптимизационни методи: (i) диференциална еволюция (Differential Evolution – DE), (ii) оптимизация с рояк от частици (Particle Swarm Optimization – PSO) и (iii) обединен метод за оптимизация с рояк от частици (Unified Particle Swarm Optimization – UPSO):

- се реализират предимно с математически операции сумиране и умножение, присъщи за реализацията на цифровите системи,
- не използват производни или други сложни нелинейни операции,
- са пригодни за изчисление с паралелни аппаратни структури,

което ги прави особено удобни за реализация в FPGA. В доклада се изследват по пет размера на популацията (10, 20, 30, 40, 50 частици) за всеки от гореспоменатите три метода, при което изследването обхваща общо 15 различни алгоритъма за обучение на PNN.

Сравнителен анализ на тези 15 алгоритъма за обучение е извършен следвайки общ експериментален протокол, който използва базата данни Parkinson Speech Dataset. Този набор от данни съдържа записи на реч от възрастни хора с болестта на Паркинсон. Речевите сигнали се обработват, сегментират и от всеки отрязък се изчисляват 26 описатели от типа на кепстралните коефициенти (Mel-frequency Cepstral Coefficients -- MFCC). Тези описатели се подават към входа на PNN, като целта е да се открият записите на реч, принадлежащи на пациенти с болестта на Паркинсон. Във всички експерименти е използвано десетократно кръстосано валидиране с разделяне на данните съответно 90% за обучение на PNN и 10% за тестване.

Анализ на експерименталните резултати показва, че най-добрият компромис между сложност при реализация и изчислителни изисквания се предоставя от методите “DE/best/1/bin” и “UPSO с  $u=0.5$ ”. Двата метода водят до сходни резултати по отношение на точност при класификация и изпреварват останалите 13 конфигурации. Въпреки че методът за обучение с DE е концептуално по-лесен за реализация е отчетено, че методът „UPSO с  $u=0.5$ “ се нуждае от по-малък брой изчислителни операции. Въпреки слабата чувствителност към грешки от закръгление на стойностите, поради реализацията на изчисленията с краен брой битове в FPGA, която е характерна за всичките три метода (DE, PSO, UPSO), авторите считат, че са необходими по-задълбочени изследвания на хардуерните реализации на тези методи, за да се оцени при кой метод се получава по-слабо понижение на общата точност, в сравнение със тази на софтуерната реализация, докладвана тук.

**[Г8.4]. Paunov P., T. Ganchev** (2014). "Low-complexity method for height estimation from speech", *Списание „Акустика“*, ISSN 1312-4897, CEEC– София, България, брой 16, Декември, 2014, стр.121-124.

Автоматизираната оценка на ръста на човек по неговия глас има за цел да предскаже височината за неизвестен от спонтанна реч. Допуснато е, че това е възможно заради наличието на силна корелация между височината на човек и дължината на речевия тракт и поради силната зависимост на характеристиките на речта от дължината на речевия тракт. Настоящото изследване е мотивирано от стремежа за интегриране на такава функционалност в речеви мобилни приложения, които се достъпват през смартфон.

В доклада [Г8.4] се изследва възможността за създаване на система за автоматизирано изчисляване на височината на непознати хора по гласа им, която да е с понижени изисквания за изчислителна мощ. За целта е извършен сравнителен анализ на четири реализации, използващи класификатори с ниска сложност по отношение на времето за изчисление и средната квадратна грешка (MSE). В изследването са анализирани точността на предсказване и времето за изчисление за (i) обобщена регресионна невронна мрежа (Generalized Regression Neural Network -- GRNN), (ii) регресия на базата на многослойна невронна мрежа от перцептрони (feed-forward Multi-layer Perceptron Neural Network -- MLP NN), регресия чрез опорни вектори (Support Vector Regression -- SVR) с ядро използвашо радиалната базисна функция (Radial Basis Function -- RBF) и с полиномиално ядро (polynomial kernel -- PK).

В изследванията е използван стандартен експериментален протокол, както е дефиниран в базата данни TIMIT, съдържаща записи на реч от 640 человека. Следвано е стандартното разделение на данните. След предварителна обработка речевият сигнал е сегментиран на кратки, при покриващи се отрезъци. От всеки отрезък са изчислени 50 описатели на речта, които формират входния вектор за регресионните алгоритми.

Анализът на експерименталните резултати показва, че когато гласовете на мъжете и жените се моделират поотделно, грешката при предсказване на височина е по-малка. Като се вземе в предвид сравнението на точността на предсказване и времето за изчисление на резултата, се оказва, че многослойна невронна мрежа от перцептрони (MLP NN) осигурява най-добрая компромис между изчислителните изисквания и точността на оценка на височината и следователно е най-подходяща за реализация в мобилни устройства.

Изследването е частично подкрепено от проект ОП "Конкурентоспособност" BG161PO003-1.2.04-0044, (2013-2014), "Развитие на устойчив капацитет на център за приложни технологии свързани със здравето", финансирали от структурните фондове на Европейския съюз и от проекта ИСП1 (2014), „Развитие на устойчив капацитет за върхови постижения в областта на биометрията, биоакустиката и биомедицинските технологии“, финансиран от Технически университет – Варна.

[Г8.5]. 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. pp. 608-616.

Автоматичното разпознаване на реч е функционалност, присъща за съвременните диалогови системи, които са неизменна част от системите за предоставяне на информационни услуги, средства за контрол и управление на различни софтуерни и хардуерни приложения - от мобилния телефон, до интелигентния дом. Голямата част от съвременните средства за разпознаване на реч все още използват стандартната комбинация от Мел-кепстрални коефициенти (Mel-frequency cepstral coefficients, MFCC), като описатели на речта и скрити модели на Марков (Hidden Markov Models, HMM) за разпознаване на речта. Тази тяхна популярност се дължи главно на широкото разпространение на софтуер с отворен код, който използва тази базова комбинация от методи, поради което тя е добре позната.

В статията [Г8.5] е представено сравнително изследване на осем различни методи за изчисляване на описатели на речта с цел оценка на пригодността им за подобряване точността на автоматичното разпознаване на реч. Освен традиционните MFCC, линейни кепстрални коефициенти (Linear Frequency Cepstral Coefficients, LFCC) и описателите Perceptual Linear Prediction (PLP CC) са изследвани и някои наскоро предложени и поради това по-малко проучени методи за параметризация на речта, изчислени посредством дискретното преобразуване на Фурье (Human Factor Cepstral Coefficients, HFCC), както и методи за параметризация използващи дискретното преобразуване с уейвлет пакети (DWPT), като (i) SBC на Sarikaya и Hansen, (ii) WPF на Farooq и Datta, (iii) WPSR на Siafarikas et al. и (iv) Overlapping WPF на Siafarikas et al.

В доклада е изведен обобщен модел на изчисляването на различните видове параметри на речта, които се различават по вида на време-честотното преобразуване, по дизайна на банката филтри, по начина, по който се представя нелинейното взаимодействие на гръмкостта на звука в човешкото ухо и др.

Оценка на практическата полезност на тези по-малко проучени методи за параметризация на речта е извършена с използването на базата данни от речеви сигнали TIMIT и система за разпознаване на реч, базирана на платформата с отворен код Sphinx-III, която предоставя среда и софтуерни инструменти за създаване на собствени системи за разпознаване на реч.

Анализ на резултатите за осемте изследвани методи за параметризация на реч показват, че всички описатели, използващи време-честотен анализ, основаващ се на дискретното преобразуване с уейвлет пакети (DWPT), довеждат до системи за разпознаване на реч с по-ниска грешка от тези с традиционните описатели MFCC и PLP CC. Конкретно SBC описателите намаляват относителната грешката от класификация на думи (WER) с 20% до 30% спрямо широкото използваните описатели PLP CC и MFCC.

Изследванията са частично подкрепени от проекта MoveOn (IST-2005-034753) „Multi-modal and multi-sensor zero-distraction interaction interface for two wheel vehicles ON the move”, съфинансиран по програмата FP6 на Европейската комисия.

[Г8.6]. F. Feradov, T. Ganchev, I. Ivanov (2015). "Detection of Sleep Disorders Based on Time-Domain EEG Signal Descriptors". Годишник на ТУ-Варна – 2015, pp.54-58.

Медицинските състояния, свързани с нарушения на съня, влияят на качеството на живот и могат да доведат до сериозни негативни ефекти върху цялостното здраве на човека. Разработването на инструменти за автоматизирано разпознаване на нарушения на съня е от съществено значение, тъй като определени нарушения на съня са симптоми на други сериозни заболявания.

В доклада [Г8.6] е представено изследване на възможността за автоматично откриване на нарушения, свързани с периодично движение на крайниците (Periodic Limb Movement Disorder - PLM) и REM Behavior Disorder (RBD) от електроенцефалографски (EEG) сигнали. За целта е изследвана възможността и целесъобразността за използване на три различни метода за машинно самообучение, като е оценена точността на класификация при откриване на нарушения на съня от типа RDB или PLM. По-конкретно за всеки от тези два вида разстройства на съня са обучени детектори с: (i) многослойна мрежа от перцентрони (MLP), (ii) класификатор с опорни вектори (SMO) и (iii) дърво на решението (J48).

Изследванията са проведени с помощта на експериментален протокол, използващ CAP Sleep Database, която съдържа 108 записи на физиологични сигнали от спящи хора. От тях 16 са на здрави хора, а останалите 92 са от пациенти с диагностицирани неврологични нарушения. В експериментите са използвани само записите на ЕЕГ сигнали за 30 человека, от които 10 здрави и 20 - с PLM или RDB нарушения. От ЕЕГ сигналите са извлечени 7 описатели, които се изчисляват директно във времевата област: *средна стойност, стандартно отклонение, кратковременна енергия, модул на средната стойност на първата разлика на сигнала, модул на средната стойност на втората разлика на сигнала, модул на средната стойност на първата на разлика на нормализирания сигнал, модул на средната стойност на втората на разлика на нормализирания сигнал*. Изчислените описатели се подлагат на статистическа нормализация до нулева средна стойност и единично стандартно отклонение. Допълнително при изчисляването на средна стойност на модула на първата и втората разлика на отчетите на нормализирания сигнал, ЕЕГ записите също се подлагат на статистическа нормализация преди изчисляването им.

Анализ на експерименталните резултати показва, че за отделните хора точността на класификация се изменя в широки граници – от 60% до 97.4%. Средната точност, постигната от различните класификатори е от порядъка на 75.2% до 76.9% за PLM, и 77.6% до 80.1% за RDB. Няма ясно изразен най-добър класификатор, т.к. класификаторът с опорни вектори SMO, който е най-точен за RDB, се явява най-неточен при PLM. Същото важи и за дървото на решението J48, но при обратен ред на точността за PLM и RDB.

Изследванията са частично подпомогнати от проекти ПД8/2015 "Разработка и усъвършенстване на методи за автоматизиран анализ на ЕЕГ сигнали, с цел откриване на негативни емоционални състояния и неврологични нарушения" и СНП2/2015 "Технологична подкрепа за подобряване качеството на живот на хора с болестта на Алцхаймер", финансираны от Технически университет - Варна.

[Г8.7]. N. Dukov, T. Ganchev (2017). “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

При традиционната архитектура на локално рекурентна, вероятностна, невронна мрежа (LRPNN), в локално рекурентния слой се използват неврони със сигмоидна активираща функция. Основна трудност при хардуерната реализация на сигмоидната активираща функция е нейната силна нелинейност, което изисква реализация с големи таблици на съответствията (look-up-tables) с много битове за всяка стойност. Това означава заделяне на големи масиви памет в FPGA чипа, което прави дизайна неподходящ за паралелна реализация.

В доклада [Г8.7] се изследват модифицирани варианти на LRPNN, при които в локално рекурентния слой се използват неврони с удобни за хардуерна реализация активиращи функции, като Rectified Linear Units (ReLU), Leaky ReLU и Clipped ReLU.

Експерименталният протокол е насочен към автоматично разпознаване на негативни емоционални състояния от ЕЕГ сигнали. За целта се използват записи на 32-канални ЕЕГ сигнали от 10 человека, които са част от базата данни с физиологични сигнали DEAP. След предварителна обработка, ЕЕГ сигнала се сегментира на кратки отрезъци. Всеки два съседни отрезъка се припокриват с 66%. От тези отрезъци се изчисляват описатели, които представляват нормализираната кратковременна енергия на сигнала за всеки от каналите. По такъв начин се формират 32 стойности на описателите за всеки зададен момент от време, които се подават на входа на LRPNN като вектор с описатели.

В изследването е определена точността при класификация на негативни емоционални състояния, при реализация на невроните в локално рекурентния слой на LRPNN със сигмоидна, ReLU, Leaky ReLU или Clipped ReLU активираща функция.

Анализ на експерименталните резултати показва, че използването на активиращи функции, базирани на Rectified Linear Units (ReLU), вместо сигмоидната функция, подобрява точността на класификация. Най-висока точност е получена при използването на Leaky ReLU, при която подобрението на точността при класификация е с приблизително 11.6%, спрямо традиционната сигмоидна активираща функция. При Clipped ReLU подобрението на точността е с 8%, а при ReLU е с 5.9%.

В допълнение на подобрените класификационни свойства на LRPNN при откриването на негативни емоционални състояния, активиращите функции от семейството на ReLU са по-лесни за изчисляване. Това прави ReLU активиращите функции подходящ избор, когато се изисква ускоряване на работата на LRPNN или хардуерна реализация в FPGA чип.

Научноизследователската и научно-приложна дейност, представени в доклада са частично подкрепени от проект НП7/2017 “Методи и средства за събиране на биомедицинска информация чрез мобилни информационни системи“, финансиран от Технически университет- Варна.

[Г8.8]. 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.

В доклада [Г8.8] се изследват възможностите за разпознаване на мисловни команди от ЕЕГ сигнали единствено чрез използване на кратковременната енергия на ЕЕГ сигнала. Разработката е насочена към развитие на интерфейси за мисловно управление на обекти (BCI – Brain-Computer Interface).

Кратковременната енергия на ЕЕГ сигнала е избрана за описател в настоящата разработка, защото позволява постигане на добро време-пространствено представяне на мозъчната активност. Кратковременната енергия се изчислява за всеки канал на многоканални ЕЕГ сигнали за кратки припокриващи се отрязъци от ЕЕГ сигнала. Отрязъците се получават чрез сегментация на ЕЕГ сигналите посредством прозоречна функция с фиксирана дължина и припокриване от 66% между съседните сегменти. Кратковременната енергия на сигнала се изчислява от всеки отрязък, като получените резултати се подлагат на статистическа нормализация. Нормализираните стойности на кратковременната енергия, изчислени за отделните канали в конкретен момент, формират вектор от описатели, който се използва в процеса на класификация.

В изследването е проведено с помощта на експериментален протокол, използващ базата данни BCI Competition IVq съдържаща записи на ЕЕГ сигнали. В частност се използва подгрупата данни „2a, aka Continuous Multi-class Motor Imagery”, включваща записи от 9 участника, които изпълняват 4 задачи, свързани с въображаемо движение на лява ръка, дясна ръка, двата крака и език. В конкретния експериментален протокол се оценява точността при идентификация на въображаеми двигателни задачи.

Идентификацията се извършва посредством класификатор с опорни вектори (SVM), използващи ядро с радиални базисни функции (rbf), обучен за всеки човек поотделно. Параметрите на SVM класификатора се настройват посредством метода за претърсване на оптимална точност по мрежа от стойности (*grid search*).

Получената средна точност за 4-те класа е в диапазона 65.6% – 71.0%. Анализът на експерименталните резултати показва, че кратковременната енергия на ЕЕГ сигналите е приложима при изготвянето на системи за мисловно управление на обекти (BCI – Brain-Computer Interface).

Невисоката класификационна точност предполага необходимостта от по-нататъшно изследване на предложения описател и комбинирането му с други описатели, с цел подобряване на точността при идентификация на двигателни дейности.

Изследванията по тематиката са подкрепени от научно-изследователските проекти PD8/2015 „Разработка и усъвършенстване на методи за автоматизиран анализ на ЕЕГ сигнали, с цел откриване на негативни емоционални състояния и неврологични нарушения“ и СНП2/2014 „Технологична подкрепа за подобряване качеството на живот на хората с болестта на Алцхаймер“, целево финансиирани от Техническия университет - Варна.

**[Г8.9].** Ф. Ферадов, Ж. Жеков, Т. Ганчев (2015). “Оценка на ефективността на статистическите описатели, извлечени от времевата област, при класификацията на епилептични пристъпи”, Сборник доклади „Иновации и Бизнес“ 2015 «Приложни технологии свързани със здравето», Окт. 9-10, ТУ-Варна

В доклада [C8.9] е представено изследване на ефективността на набор от седем статистически описатели на ЕЕГ сигнали за нуждите на автоматично откриване на епилептични пристъпи. Епилепсията е широко разпространено неврологично заболяване, водещо до внезапни епизоди на нарушение на двигателните функции, загуба на съзнание и конвулсии. Поради тази причина откриването и прогнозирането на епилептични пристъпи е проблем с голяма социална значимост.

Избранныте седем статистически описатели на ЕЕГ сигнали се извличат директно във времевата област, използвайки отчетите на ЕЕГ сигналите, посредством прозорец с фиксирана дължина. При покриването между съседните сегменти е 66%. За всеки времеви отрезък, получен след сегментиране на ЕЕГ сигнала, се изчисляват седем статистически описатели: *средна стойност, стандартно отклонение, кратковременна енергия, модул на средната стойност на първата разлика на сигнала, модул на средната стойност на втората разлика на сигнала, модул на средната стойност на първата на разлика на нормализирания сигнал, модул на средната стойност на втората на разлика на нормализирания сигнал*, изчислени директно във времевата област.

Експерименталният протокол използва базата данни от записи на ЕЕГ сигнали на трима пациенти с епилепсия в различно здравословно състояние, изготвена съвместно между университетите в Берн и Барселона. ЕЕГ сигналите се разделят на 100 сегмента с продължителност 23.6 сек, като общата им продължителност е 40 мин. След изчислението на описателите, сигналите от тримата пациенти са обединени в един общ масив, съдържащ 300 сегмента.

Получените масиви от описатели са използвани за създаването модели за автоматична класификация на епилепсия от ЕЕГ записи. Използвани са три метода за класификация: многослойна невронна мрежа от перцептрони (MLP NN), класifikатор с опорни вектори (SMO), дърво на решенията (J48). Средната точност при откриване на епилептични пристъпи е в диапазона от 98.7% до 100%. Високият резултат при класификация е индикация за качествата на избрания набор от седем статистически описатели на ЕЕГ сигнала.

Изследванията по тематиката са подкрепени от научно-изследователските проекти PD8/2015 „Разработка и усъвършенстване на методи за автоматизиран анализ на ЕЕГ сигнали с цел откриване на негативни емоционални състояния и неврологични нарушения.“ и СНП2/2014 „Технологична подкрепа за подобряване качеството на живот на хората с болестта на Алцхаймер“, целево финансиирани от Техническия университет - Варна.

[Г8.10]. N. Dukov, T. Ganchev, D. Kovachev (2015). "Reduced-complexity FPGA-implementation of PNN for classification of binary images". Годишник на ТУ-Варна – 2015, pp.59-63.

Вероятностната невронна мрежа (PNN) намира широко приложение заради способността си да създава устойчиви класификатори при минимален набор от образци за всяка категория данни. Обучението на PNN се заключава в запомняне на набора от образци и настройка на един или няколко параметри, които определят степента на гладкост на функцията на разпределение на вероятността за всяка категория данни. Възможността по-голямата част от изчисленията да се извършват паралелно прави PNN особено удобна за хардуерна реализация с FPGA чип.

В доклада [Г8.10] се изследват качествата на хардуерна реализация на PNN, която е създадена и оптимизирана специално за класификация на бинарни изображения, с цел откриване на сърдечни заболявания. Хардуерната реализация използва ресурсите на чип Altera Cyclone IV. Проведени са изследвания доколко някои опростявания на невроните в образцовия слой на PNN повлияват на точността при класификация.

Изследванията са проведени чрез експериментален протокол, използващ базата данни SPECT (Single Proton Emission Computed Tomography images). SPECT съдържа томографски снимки на сърцата на 267 пациенти, част от които са диагностицирани със сърдечни заболявания. За всяка снимка от базата данни са извлечени 22 бинарни описатели, характеризиращи изображението, които се обединяват в един вектор и се подават на входа на PNN. В случая PNN е използвана за създаване на детектор на миокардна перфузия. Изследването обхваща сравнение на софтуерната и хардуерната реализация на PNN с топология: 22 входа, 50 неврона в образцовия слой, два неврона в сумирация слой и един неврон в изходния слой. Реализирани са PNN с и без отместване (*bias*) в невроните от образцовия слой и за две различни стойности на отместването, когато *bias* е реализирано. Анализът на експерименталните резултати показва, че:

- Точността на класификация при софтуерната реализацията в среда на MATLAB и представяне на стойностите с плаваща запетая и 64-битова прецизност, превъзхожда апаратната реализация с 0.5%-1.5%, в зависимост от това, дали е реализирано отместване или не.
- Реализацията на отместване (*bias*) позволява подобряване на точността при класификация и за софтуерната и за хардуерната реализация на PNN с 0.5%-1%. Оптималната стойност на отместването (*bias*) е различна при софтуерната и хардуерната реализация на PNN.
- При използване на оптималните стойности на отместването (*bias*), разликата в точността на класификация между софтуерната и хардуерната реализация на PNN е само 1%. В конкретния дизайн това означава, че могат да бъдат икономисани между 24 000 и 36 000 логически елемент от ресурсите на FPGA чипа без значителна загуба на точност.

Изследванията са частично подпомогнати от проект ПД5/2015 “Изследване на биологично мотивирани архитектури на невронни мрежи за разпознаване на сърдечни заболявания и неврологични разстройства”, СНП2/2015 “Технологична подкрепа за подобряване качеството на живот на хора с Алцхаймер” финансираны от ТУ - Варна.

**[Г8.11]. 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.

Електроенцефалографията дава възможност за пряко наблюдение и идентифициране на емоциите в момент на тяхното възникване. Сред основните предизвикателства пред автоматизираното разпознаване на емоции е създаване на описатели, които представлят характеристиките на ЕЕГ сигнала, носещи ясни и недвусмислени доказателства за проявените емоции, докато остават нечувствителни към изменчивостта, дължаща се на друга мозъчна активност.

В доклада [Г8.11] се представят резултати от експерименталната оценка на целесъобразността на седем описатели на ЕЕГ сигнала, изчислени директно във времевата област. Основното предимство на тези описатели е ниската сложност на тяхното изчисление, което ги прави подходящи за използване в системи с ограничен ресурс и ниска консумация на енергия.

Изследването съпоставя *кратковременната енергия* на ЕЕГ сигналите с широко използван набор от шест статистически описатели включващ: *средна стойност, стандартно отклонение, модул на средната стойност на първата разлика на сигнала, модул на средната стойност на втората разлика на сигнала, модул на средната стойност на първата на разлика на нормализирания сигнал, модул на средната стойност на втората на разлика на нормализирания сигнал*. Избраните описатели се изчисляват за кратки отрезъци от ЕЕГ сигналите, получени посредством правоъгълен прозорец с фиксирана дължина. Процедурата се повтаря за всеки канал на ЕЕГ записа, като получените описатели се подлагат на статистическа нормализация. Изследвани са три различни набора, съдържащи различен брой описатели: (i) самостоятелно използване на кратковременната енергия, (ii) използване на комбинация от шестте статистически описатели и (iii) едновременно използване на всички 7 статистически описатели.

В експерименталното изследване са използвани ЕЕГ записи от базата данни DEAP. За всеки участник от базата данни се създава класификатор с опорни вектори (SVM), използваш ядро с радиални базисни функции (rbf). Настройката на параметрите на SVM класификаторите се извършва чрез претърсване на стойностите на параметрите “boxconstraint” и “rbf sigma” по мрежа от предефинирани стойности (*grid-search*).

Средна точност, получена в резултат на експерименталната оценка на описателите, варира в диапазона 65.7% - 67.8%, като се наблюдава малка разликата в точността между отделните групи описатели. Анализ на резултатите показва, че разликата между средната точност за набора от седем статистически описатели и самостоятелното използване на кратковременната енергия е 2.1%, а разликата на средната точност между наборите от шест и седем описатели е 0.4%. Отчита се широк диапазон на изменчивост на точността за различните участници – до 33%. Докладваме, че комбинираното използване на всичките седем описатели, води до известно подобряване на точността на разпознаване в сравнение с подгрупите описатели, изследвани в предишни разработки.

Изследванията са подкрепени от проектите по ОП „Конкурентоспособност“ BG161PO0031-1.2.04-0044 и ОП „РЧР“ BG051PO001-3.3.06-0005 финансиирани от структурните фондове на ЕС, както и от проекта ISP1/2014 финансиран от Технически университет - Варна.

**[Г8.12]. 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

В доклада [Г8.12] е представено изследване на индивидуалната полезност на 10 статистически описатели на ЕЕГ сигналите за нуждите на разпознаване на негативни емоционални състояния. Измежду тези често използвани статистически описатели на сигналите, изчислявани от времевата област, са *средна стойност, стандартно отклонение, кратковременна енергия, модул на средната стойност на първата разлика на сигнала, модул на средната стойност на втората разлика на сигнала, модул на средната стойност на първата на разлика на нормализирания сигнал, модул на средната стойност на втората на разлика на нормализирания сигнал, активност, мобилност и комплексност*. Резултатите от изследването са част от усилията за намиране на компактен набор от полезни описатели, с който да се понижи общия обем на изчисленията, без да се намалява прецизността при разпознаване на емоционални състояния от ЕЕГ сигнали.

Експериментална постановка използва ЕЕГ сигнали от базата данни DEAP, като изследването включва всички ЕЕГ записи, предоставени в базата, като отпадат единствено участниците, при които след разпределение на записите един от двата класа съдържа по-малко от 20% от общия брой данни. Данните на участниците се разделят на две категории – „негативни“ и „други“. Разделянето се извършва въз основа на критериите за оценка на стимулите: „възбуда“, „доминантност“, „валентност“ и „харесване“. За всички параметри записите, получили оценка по-голяма от 4 се класифицират като „други“, докато записите с оценка по-ниска от 4, се класифицират като „негативни“.

За определяне на значимостта на описателите е използван алгоритъма *ReliefFAttributeEval*, като се оценява корелацията на всеки описател спрямо всеки от изследваните класове и се генерира списък, в който описателите са класирани по значимост в низходящ ред. Получените списъци се обобщават за всички участници и се провежда цялостна оценка и анализ на получените резултати.

Анализ на експерименталните резултати показва, че статистическите описатели от по-висок ред, като *модул на средната стойност на първата на разлика на нормализирания сигнал и комплексност*, имат по-голяма статистическа значимост в сравнение с описателите от по-нисък ред. Като цяло описател с най-високо ниво на корелация спрямо емоционалните състояния в ЕЕГ сигнали е *модула на средната стойност на втората производна на нормализирания сигнал*. Най-ниска корелация се наблюдава при описателите *активност и кратковременна енергия*.

Анализът на резултатите показва, че макар и да се наблюдават някои различия, критерият използван за разделяне на данните, не оказва значително влияние върху дискриминативните свойства на описателите в тези близко свързани задачи. В същото време обаче, се наблюдават големи разлики между описателите, избрани за отделните хора. За да се реши този проблем е необходимо системата за разпознаване на емоционални състояния да изчислява голям набор от описатели и да е в състояние автоматично да подбира подходящото подмножество, които са полезни във всеки конкретен случай.

[Г8.13]. P. Zervas, T. Ganchev, N. Fakotakis (2006). "Negative Emotional State Detection from Speech", *Proc. of the Proc. of the Communication Systems Networks and Digital Signal Processing*, CSNDSP-06, 19-21 July, 2006. pp. 310-313.

Откриването на негативни емоционални състояния е от голямо значение в интерфейсите човек-машина. В частност това е в сила и приложениета от типа: (i) интелигентен дом, (ii) интелигентни речеви интерфейси за управление и контрол на роботи или (iii) информационни услуги осигуряващи достъп до речеви диалогови системи.

Разработката е насочена към създаване на диалогова система за управление на интелигентен дом чрез реч, включително и по телефона. Способността на диалоговата система да разпознава негативни емоционални състояния, разширява функционалността с цел, да се получи информация за конфликтна ситуация, разочарование или друго негативно състояние на потребителя. Такава функционалност позволява да се смени стратегията за водене на диалога, да се направи профилиране на диалога според потребителя и др.

В доклада [Г8.13] се изследват възможностите за разпознаване на четири вида негативни емоционални състояния от говорима реч: „горещ гняв“, „хладен гняв“, „презрение“, „отвращение“, както и неутрално емоционално състояние. Извършен е анализ на същността на тези емоционални състояния и са създадени акустични модели и класификатор, който да ги разграничава. За целта е изследвана приложимостта на два класификатора: вероятностни невронни мрежи (PNN) и дърво на решенията C4.5.

За нуждите на изследванията е дефиниран общ експериментален протокол, който използва базата данни от записи на емоционална реч, LDC Emotional Prosody Speech, разработена от Университета на Пенсилвания. Базата данни съдържа 9 часа емоционална реч, записана от актьори и съответните анотации на записите. От базата данни са използвани само записите с речеви сигнали, които принадлежат към петте гореспоменати категории: горещ гняв, хладен гняв, презрение, отвращение и неутрално емоционално състояние.

След предварителна обработка и сегментация от речевите сигнали са формирани кратки отрязъци с дължина 16 милисекунди, които се припокриват с 50%. За всеки отрязък са изчислени 19 описатели, характеризиращи прозодични параметри на речта (период на основния тон, енергия и техните производни), формантите на речевия тракт (F1, F2, и F3), и кепстрални параметри (MFCC). Така са формирани входни вектори, съставени от акустични атрибути, свързани с прозодия и конфигурация на речевия тракт, които се подават към класификаторите.

Анализът на експерименталните резултати показва, че вероятностната невронна мрежа PNN превъзхожда значително дървото на решенията C4.5 в конкретната задача. Общата точност за PNN при класификация на петте категории емоции е 81.1%, докато при C4.5 точността е 62.8%.

Разглеждайки точността на класификация, постигната за отделните категории емоции се оказва, че единствено за категория „студен гняв“ класификаторът с дърво на решенията C4.5 показва по-добри резултати от PNN, но за всички останали категории PNN превъзхожда дървото на решенията.

[Г8.14]. Dukov N., Ganchev T., Kovachev D. (2014). "Empirical Study of PNN Sensitivity to the Exponential Function Approximation in FPGA Implementations", *Годишник на ТУ-Варна* -- 2014. vol.1, pp.53-58.

Хардуерната реализация на алгоритмите за машинно обучение често изисква изчисляване на нелинейни функции, като например експоненциална функция, естествен логаритъм, синус и косинус и др. В частност при реализация на Вероятностната невронна мрежа (PNN) е необходимо да се реализира експоненциалната активираща функция на невроните от образцовия слой, която най-лесно се реализира с таблица на съответствията (look-up-table).

В доклада [Г8.14] се изследва влиянието на точността на апроксимация на активираща експоненциална функция на невроните от образцовия слой върху точността при класификация на изображения. Реализирани са апроксимации с различни размери на таблица на съответствията, които реализират различно приближение до идеалната експоненциална функция и са изследвани ефектите, свързани с ограничната точност на представяне на числата.

Използвайки пет апроксимации на експоненциалната функция в хардуерна реализация на детектор на сърдечни заболявания, на базата на вероятностни невронни мрежи (PNN) е проведено експериментално изследване за автоматично откриване на миокардна перфузия в томографски изображения на сърдечната област. Хардуерната реализация използва ресурсите на чип Altera Cyclone IV. Изследванията са проведени чрез експериментален протокол, използващ базата данни SPECT (Single Proton Emission Computed Tomography images). SPECT съдържа томографски снимки на сърцата на 267 пациенти, част от които са диагностицирани със сърдечни заболявания. За всяка снимка от базата данни са извлечени 22 бинарни описатели, характеризиращи изображението, които се обединяват в един вектор и се подават на входа на PNN с топологии: 22 входа, 80 (или 50) неврона в образцовия слой, два неврона в сумирация слой и един неврон в изходния слой.

Анализ на експерименталните резултати показва, че при 80 неврона в образцовия слой, точността при класификация за хардуерната реализация на PNN отстъпва на софтуерната само с 0.6%. При 50 неврона разликата надхвърля 2%, главно заради големия брой случаи, при които PNN изчислява еднаква вероятност за двете категории. Стъпката, с която се реализира LUT апроксимацията на експоненциалната функция, не оказва значително влияние върху точността на класификация при 80 неврона в образцовия слой. Времето за изчисления на резултата при класификация на един входен вектор не се променя при различна прецизност на апроксимация на LUT. При реализация на PNN с 50 неврона, изходният резултат се изчислява за 10 цикъла на тактовия сигнал, а при реализация с 80 неврона в образцовия слой за 7 цикъла.

Научно-изследователската и развойна дейност, свързана с това изследване е проведена с логистична подкрепа от проект ОП "Конкурентоспособност" BG161PO003-1.2.04-0044, (2013-2014), "Развитие на устойчив капацитет на център за приложни технологии свързани със здравето", финансиран от Структурни фондове на Европейския съюз и от проект ИСП1/2014 „Развитие на устойчив капацитет за върхови постижения в областта на биометрията, биоакустиката и биомедицинските технологии“, финансиран от Технически университет - Варна.

**[Г8.15]. A. Lazaridis, I. Mporas, T. Ganchev** (2012). “Phone Duration Modeling of Affective Speech using Support Vector Regression”, *International Journal of Intelligent Systems and Applications* (IJISA), ISSN: 2074-904X (Print), ISSN: 2074-9058 (Online), vol.4, no.8, 2012. pp.1-9.

Качественото моделиране на продължителността на фонемите е предпоставка за синтезиране на емоционална реч с естествено звучене. Основен акцент на изследването, представено в статията [Г8.15] е оценка на ефективността при моделиране и предсказване на продължителността на фонемите чрез регресия с опорни вектори (Support Vector Regression -- SVR) в контекста на емоционалната реч. Това е новост, тъй като към момента на публикуване на статията SVR не е бил използван за моделиране на продължителността на фонемите в контекста на емоционална реч.

В допълнение, в статията се извършва сравнителен анализ с други десет метода за моделиране и предсказване на продължителността на фонемите. Тези модели принадлежат към добре известни и широко използвани категории алгоритми за машинно самообучение, сред които са дървета на решениета, линейна регресия, методи с мързеливо обучение и алгоритми базирани на мета-обучение.

В изследването е използван общ експериментален протокол за оценката на единадесетте метода за предсказване на продължителността на фонемите. За целта е използвана база данни от записи на емоционална реч на съвременен гръцки език. Базата данни съдържа изрази и фрази произнесени от професионален актьор по пет различни начина -- четири са в категорията емоционална реч (гняв, страх, радост, тъга) и петият е неутрална реч. За всеки запис са изчислени 93 описатели на речта, които формират входния вектор към единадесетте метода за регресия изследвани тук.

Анализ на експерименталните резултати показва, че моделирането на продължителността на фонемите с SVR превъзхожда останалите десет модела във всичките четири категории емоции. По-конкретно, по отношение на критерия RMSE (корен квадратен от средно квадратичната грешка), SVR моделът постига средно относително намаление на грешката от предсказване на продължителността с 8% във всички емоционални категории. Относителното понижение на RMSE за отделните категории реч е от 6.8% до 9.2% в сравнение с втория най-добър модел.