РЕЗЮМЕТА

НА НАУЧНИ ТРУДОВЕ И УЧЕБНИ ПОСОБИЯ

на ас. д-р инж. Тодор Димитров Ганчев

представени за участие в конкурса за доцент по професионално направление: 5.2. „Електротехника, електроника и автоматика“, учебна дисциплина: „Цифрова обработка на сигналите“, публикуван в ДВ брой 48 / 31.05.2013 г.

За участие в конкурса са подбрани **общо 42 рецензирани публикации**, в т.ч. **40 научни публикации** и **2 учебни пособия**, разделени по категории както следва:

- Монографичен труд на английски език (издание на Springer, New York), 1 бр.;
- Статии в международни научни списания в чужбина (издания на IEEE, Elsevier, Springer, и др.), 12 бр.;
- Глави в книги от международни научни издателства (издания на Springer, IGI-Global и др.), 12 бр.;
- Доклади в сборници от международни конференции в чужбина, 15 бр.;
- Учебни пособия (издателство ТУ-Варна), 2 бр..

От представените за участие в конкурса общо 40 публикации, **30 са поместени в база данни Scopus**. Десет от научните трудове представени в конкурса са публикувани в **списания с импакт фактор** (**ISI Thomson Reuters Impact Factor)**, **и имат кумулативен импакт фактор IF2012 = 11,427**. Подробна справка за публикациите с импакт фактор е предоставена в раздел *Списък на публикациите с импакт фактор* на стр. 7 в документите по точка 5. *Списък на научните трудове и учебни пособия*.

Научните публикации представени за участие в конкурса за разделени в **две групи**.

Първата група (А), общо **20 публикации**, **са обединени като равностойни на монографичен труд** на тема **"ЦИФРОВА ОБРАБОТКА НА РЕЧЕВИ СИГНАЛИ":**

- Монографичен труд на английски език 1 бр.;
- Статии в *международни научни списания в чужбина*, 6 бр.;
- Глави в *книги от международни научни издателства*, 6 бр.;
- Доклади в сборници от *международни конференции в чужбина*, 7 бр..

Тематично трудовете от **група А** са систематизирани в следните области:

1. Нови методи за извличане на дескриптори на речевия сигнал за нуждите на приложения изискващи разпознаване на реч, разпознаване на диктори, сегментиране на реч и др., (общо 6 научни публикации);
2. Нови методи за автоматична сегментация на речеви сигнали до фонеми и срички, (общо 2 публикации);
3. Нови методи за извличане на лингвистична информация от речеви сигнали -- методи за шумоустойчиво разпознаване на реч, (общо 4 научни публикации);
4. Нови методи за извличане на не-лингвистична информация от речеви сигнали -- разпознаване на диктори, разпознаване на емоционалното състояние на диктора, определяне на биометрични данни, като ръст, диалект и др., (общо 8 публикации).

Втората **група (Б)** включва **20 публикации** разпределени както следва:

- Статии в международни научни списания в чужбина, 6 бр.;
- Глави в книги от международни научни издателства, 6 бр.;
- Доклади в сборници от международни конференции в чужбина, 8 бр.;

Тематично научните трудове от **група Б** са систематизирани в следните области:

1. Усъвършенствани методи за машинно разпознаване на сигнали, (общо 7 публикации)
2. Нови методи за биоакустично разпознаване на биологичните видове по техните акустични емисии, (общо 8 публикации).
3. Методи за обработка на аудио сигнали, създаване на бази данни аудио и реч, (общо 3 публикации)
4. Методи за оценка качеството на системи за трансформация на реч (voice conversion), (общо 2 публикации)

Трета **група (В)** включва **2 учебни пособия**, издания на ТУ-Варна**.**

А. МОНОГРАФИЧЕН ТРУД И НАУЧНИ ПУБЛИКАЦИИ, РАВНОСТОЙНИ НА МОНОГРАФИЧЕН ТРУД НА ТЕМА "ЦИФРОВА ОБРАБОТКА НА РЕЧЕВИ СИГНАЛИ":

I. Монографичен труд, 1 брой

[A1] Ganchev T.: *Contemporary methods for speech parameterization*. First Edition. Springer, New York, 2011. ISBN: 978-1-4419-8446-3, DOI: 10.1007/978-1-4419-8447-0.

Извличането на дескриптори, описващи речевия сигнал, е важен етап в съвременните технологии за обработка на реч, използвани в мобилните и наземни комуникации, банкиране по телефона, автоматизирани информационни услуги и др. Основен проблем при извличането на дескриптори на речта е улавянето на полезната информация в речевия сигнал и елиминиране на нежеланите вариации, като тези са специфични според целите на приложението и средата в която се намира дикторът.

- Монографичният труд обобщава основните концепции залегнали в параметризацията на речеви сигнали, разглежда историческите корени на тези концепции и в последствие систематизира най-съвременните методи за извличане на дескриптори на речевия сигнал. Изхождайки от цялостният процес на обработка, на който се основават технологиите за обработка на реч, и като се отчитат характеристиките на речевите сигнали в телефонните комуникации, както и изискванията на съвременните стохастични алгоритми за класификация и за компресия на сигнали, се разглеждат шест метода за извличане на дескриптори, основаващи се на дискретното преобразуване на Фурие, и пет метода основаващи се на дискретното преобразуване с уейвлет пакети. Унифицираното представяне на тези методи и предложеният сравнителен анализ на техните прилики и различия значително подпомагат разбирането им и създават благоприятни условия за по нататъшното им усъвършенстване. Разгледаните методи за извличане на дескриптори на речевия сигнал са съпоставени и сравнени в три типа приложни задачи: разпознаване на фонеми, разпознаване на реч и разпознаване на диктори. Осъществен е и емпиричен сравнителен анализ на изследваните методи за извличане на дескриптори на речевия сигнал и са представени заключения за полезността на тези методи за всеки от трите типа приложни задачи.
- Централно място в книгата заема Глава 4, която представя съвременните методи за

извличане на дескриптори на речевия сигнал в унифициран вид и предлага сравнителен анализ на техните особености, прилики и разлики, но всяка от другите глави също предлага приноси и съпътстващи анализи, които допринасят за по-пълното разбиране на областта на приложимост и полезност на всеки от изследваните методи.

Накратко, монографичният труд включва следните глави:

Глава 1: Introduction - представен е цялостният процес за цифровата обработка на речевите сигнали, включващ етапа на извличане на дескриптори на речевия сигнал и предшестващите и последващи обработки. Изясняват се историческите корени на концепциите залегнали в съвременните методи за извличане на дескриптори и съвременните разбирания за критичните ленти на слуховото възприятие. Предлага се кратък исторически обзор на най-разпространените методи за извличане на дескриптори, съвременните достижения и направлението в което се очаква да продължи тяхното развитие.

Глава 2: Speech pre-processing - формулират се основните положения и етапи на цифровата обработка на сигнала, предшестваща извличането на кепстрални дескриптори на речевият сигнал.

Глава 3: DFT-based speech parameterization - въвежда се унифицирано описание на шест традиционни метода за извличане на дескриптори на речевия сигнал (LFCC, MFCC-FB20, HTK MFCC-FB24, MFCC-FB40, HFCC-E-FB29, PLP-FB19) основаващи се на дискретното преобразуване на Фурие и се извършва сравнителен анализ на техните особености, прилики и разлики.

Глава 4: DWPT-based speech parameterization - въвежда се унифицирано описание на пет съвременни метода за извличане на дескриптори на речевия сигнал (WPF-SBC, WPF-FD, WPF-OBJ, WPF-OVL, WPF-ACE) основани на дискретното преобразуване с уейвлет пакети. Предлага се сравнителен анализ на техните особености, прилики и разлики.

Глава 5: Evaluation on the monophone recognition task - представя резултати от емпирично оценяване на пригодността на 12 метода за извличане на дескриптори на речта за нуждите на задачата за разпознаване на фонеми. Извършен е задълбочен анализ на резултатите по фонеми и класове фонеми.

Глава 6: Evaluation on the speech recognition task - представя резултати от емпирично оценяване на пригодността на 11 метода за извличане на дескриптори на речта за нуждите на задачата за разпознаване на реч. Извършен е задълбочен анализ на резултатите.

Глава 7: Evaluation on the speaker verification task - представя емпирично изследване на пригодността на 10 метода за извличане на дескриптори на речта за нуждите на задачата за разпознаване на диктори. Представен е сравнителен анализ на резултати за общо 60 набора дескриптори, включващ основните дескриптори и техни подмножества.

Глава 8: Conclusion and outlook - представя цялостен поглед върху резултатите получени в Глави 4, 5, 6 и очертава областта на приложимост на отделните методи за извличане на дескрипторите на речевия сигнал. Направено е предположение за възможно бъдещо развитие на методите за параметеризация на реч в посока на доближаване към биологично мотивирани стратегии за обработка и вероятно до отпадане на понастоящем въздесъщото декорелиращо косинусово преобразуване, използвано при всички съвременни методи.

Глава 9: Links to code and further sources of information - представя информация и връзки към програми и други ресурси, които биха били полезни за младите научни работници, докторанти и студенти, които имат желание да се занимават с обработка на речеви сигнали.

Приложение: Appendix I. Other versions of the HTK MFCC filter-bank - представя две алтернативни дефиниции на банките филтри използвани за определяне на дескриптори по метода известен като **HTK MFCC**.

Статии в научни списания, общо 6 броя

[A2]. **T. Ganchev**, M. Siafarikas, I. Mporas, Ts. Stoyanova: *Wavelet basis selection for enhanced speech parametrization in speaker verification*, International Journal of Speech Technology, ISSN: 1381-2416. May 2013.

Въпреки че кепстралните коефициенти получени с използването на банка филтри базирана на Мел-скалата (Mel-Frequency Cepstral Coefficients - MFCC) станаха широко разпространени след въвеждането на стандарт от Европейският институт за телекомуникационни стандарти (European Telecommunications Standards Institute - ETSI), в съвременната научна литература има множество свидетелства, че параметризацията на речеви сигнали, базирана на преобразуване с уейвлет пакети, води до по-добри резултати в определени приложения (разпознаване на реч, диктори и др.) В същото време, обща слабост на всички предишни публикации съобщаващи за токова предимство на методите за извличане на дескриптори на речта базирани на преобразуването с уейвлет пакети е, че: (а) свойствата на избраните базисни уейвлет функции не са в фокуса на дискусията и (б) като цяло липсва дискусия относно критериите използвани при избора на базисни функции. Вместо това авторите предлагат някаква предварително определена базисна функция: (i) след експериментиране с набор от уейвлет функции от даден вид, (ii) разчитат на опит придобит при работа по подобни задачи, или (iii) са водени от вътрешни убеждения и интуиция.

- В статията се предлага по-общ подход, който позволява изборът на базисна функция за преобразуването с уейвлет пакети да бъде извършен посредством обективен метод, който се основава на теоретичен анализ на свойствата на отделните функции. Работоспособността на предложения подход е илюстрирана за задачата за разпознаване на диктори, където посредством избор на най-подходящата базисна уейвлет функция измежду 9 кандидата е постигнато намаляване на грешката при разпознаване на дикторите.
- Практическата стойност на предложеният оптимизиран метод за извличане на дескриптори на речта, който води до подобряване на точността при разпознаване на диктори, е демонстрирана на две различни световноизвестни специализирани бази данни за тестване на технологии за разпознаване на диктори. Резултатите с използването на база данни 2001 NIST Speaker Recognition Evaluation Database, състояща се предимно от записи на телефонни разговори по мобилните мрежи в САЩ, показват предимство на предложения метод с 4.2 %. Аналогично, резултатите с използването на Polycost Speaker Recognition Database, състояща се от записи на телефонни разговори по наземните телефонни мрежи на Европа, показват предимство от 2.3 % спрямо стандартизираните MFCC дескриптори, базиращи се на дискретното преобразуване на Фурие.
- Експерименталните данни потвърждават че базисната уейвлет функция на Battle-Lemarie от пети ред е най-подходящата измежду всичките девет кандидата, представители на различни уейвлет фамилии.

Експерименталните изследвания докладвани в статията са извършени следвайки стандартизиран експериментален протокол на Националният институт за стандартизация и технологии (NIST) на САЩ и използвайки система за разпознаване на диктори разработена от Т. Ганчев. Тази система е участвала в три световни състезания за оценка на технологии за разпознаване на диктори организирани от NIST, а също и в световно състезание организирано от Холандския институт по правосъдие (NFI) и Холандската организация за приложни научни изследвания (TNO), като най-доброто крайно класиране е второ място на състезанието организирано от NFI/TNO. Използването на стандартизирани база данни, експериментален протокол и система за разпознаване на диктори с известна точност създава възможност за директни сравнения с предшестващи и следващи разработки.

[А3]. I. Mporas, **T. Ganchev**, N. Fakotakis: *Speech Segmentation using Regression Fusion of Boundary Predictions*, Computer Speech & Language, ISSN: 0885-2308, vol. 24, no. 2, 2010. Elsevier, pp. 273–288. **(ISI Thomson IF 2012=1.463)**

Съвременната технология за обработка на реч широко използва статистически методи, чиято прецизност зависи в значителна степен от наличието на големи бази данни реч и съответните анотации (транскрипция по думи, прозодична анотация, транскрипция по фонеми, определяне на границите на фонемите и др.). Тези анотации най-често се определят ръчно или полуавтоматично, като при разработването на системи за разпознаване на реч (ASR) е достатъчна фонетичната транскрипция, докато за нуждите на синтеза на реч (TTS) са нужни също и границите на фонемите. Най-общо, транскрипция по думи или по фонеми е сравнително лесна задача, която се извършва автоматично и след това се проверява ръчно. В същото време, определянето на границите на фонемите се счита за трудна задача, която изисква участието на експерти по фонетика. Последното е скъпа и трудоемка процедура.

- В статията се предлага метод за комбиниране на множество предсказвания на границите на сегментите, който се основава на регресивен анализ на предсказанията получени от разнородни модели за сегментация. Методът е приложим независимо от броя на индивидуалните предсказатели и от конкретната им реализация. Същността на метода се състои в това, че регресивният алгоритъм за комбиниране на тези предсказания компенсира систематичните грешки на всеки един от индивидуалните модели и систематичните измествания на оценката за границите на сегментите за всеки тип граница (например преходите между границите на гласни, носови съгласни, фрикативни съгласни и т.н.).
- Работоспособността на метода се демонстрира при комбиниране на резултатите от голям брой (112) модели за сегментация, които се различават по използваните дескриптори на речта и по сложността на модела (различен брой състояния на скрития Марковски модел, смеси от различен брой Гаусови функции за всяко състояние, използване на моделиране с и без отчитане на контекста). Предсказанията от тези 112 модела се използват като входни данни за предложения метод за комбиниране на резултатите и за формиране на по-прецизни граници на фонемите.
- Експерименталните резултати получени при работа с база данни TIMIT показват, че комбинация на предсказанията чрез регресия с опорни вектори (SVM regression) е способна да постигне по-прецизни предсказания, в сравнение с други реализации с линейна регресия докладвани в научната литература. Демонстрирано е значително подобрение на прецизността на предсказване спрямо ней-добрият индивидуален предсказател, като за най-широко използвания толеранс от 20 ms точността е подобрена с приблизително 9 % в абсолютни единици. В допълнение средната абсолютна грешка е намалена с приблизително 33 %, а средната квадратичната грешка е намалена с 27 %.
- Методът работи и при използването на други алгоритми за регресивен анализ, но подобрението на точността е в по-малки граници. Сравнителни експерименти в еднакви условия показват че предложения метод с SVM regression дава по-прецизни предсказания в сравнение с други съвременни методи.

Експериментите са извършени на добре познатата база данни, TIMIT при прилагането на широко използван експериментален протокол. Използвано е моделиране със скрити Марковски модели, което създава предпоставки за добро комуникиране на резултатите и за директни сравнения с предшестващи и следващи разработки.

[A4]. I. Mporas, **T. Ganchev**, N. Fakotakis: *Phonetic Segmentation of Emotional Speech with HMM-based Methods*, International Journal of Pattern Recognition and Artificial Intelligence, ISSN 0218-0014, 2010. **(ISI Thomson IF 2012=0.562)**

Поради несъвършенствата на съвременните интерфейси човек-машина, и главно поради отсъствието на емоционална интелигентност при машините, по време на общуването между хора и машини често се проявява сериозно разминаване между информацията съобщавана от човека и информацията възприета от машината. За постигането на устойчива и приятна за потребителя комуникация е необходимо машините да възприемат и изразяват прецизно лингвистичната и пара-лингвистичните компоненти на съобщението, включително емоциите. Правилното възприемане на речта изисква интерфейса на машината да е снабден с модул за автоматично разпознаване на реч, който работи съвместно с модул за разпознаване на емоционалното състояние на диктора. Правилното изразяване на емоции изисква модул за синтез, който да пресъздава емоционална реч. Създаването на модули, които могат правилно да разпознават емоционална реч, се улеснява от наличието на големи бази данни и съответните анотации. От друга страна, създаването на модули за синтез на емоционална реч задължително изисква да са известни и границите на фонемите или сричките, поради което е необходимо да се извършва сегментация на големи бази данни.

- Статията разглежда проблема за сегментиране на емоционална реч, като се фокусира на методите които не се нуждаят от начална порция анотирани данни (bootstrap data). Изследва се приложимостта на класически и сравнително нови методи използващи моделиране със скрити модели на Марков (HMM), които са доработени за спецификите при сегментиране на емоционална реч. Показано е, че хибридните методи използващи вградено-изолирано обучение са по-успешни в сравнение с други широко използвани понастоящем методи, базирани на скрити модели на Марков. Подобрената прецизност на сегментация се дължи на специалното итеративното обучение, при което хибридния метод се изразява в последователно рафиниране на изчислените на предишните итерации граници на фонемите.
- Изследвана е модификация на хибридния метод, при който се създават специализирани модели за всяка отделна категория емоционална реч, за разлика от оригиналния метод при който се използва един общ модел. Експериментално е показано, че използването на усъвършенстваният метод, който използва специализирани модели за всяка категория емоционална реч, позволява подобряване на прецизността на сегментация спрямо общия случай, когато се използва един общ модел, изграден от емоционално неутрална реч. Този метод е най-успешен в сравнение с другите методи изследвани в статията.
- При толеранс за грешката ≤20 ms, подобрението на прецизността дължащо се на новия метод е приблизително 7.6 % за категория "гняв", 19.1 % за "страх", 18.5 % за "радост", 14.2 % за "тъжен" и 8.1 % за емоционално неутрална реч, в сравнение с втория най-добър метод.
- Изследван е вариант, при който сегментирането на емоционална реч се извършва с модел който е обучен за друга, несъвпадаща, категория емоционална реч. Получените резултати показват, че модели обучени с бърза реч (например "радост") се справят по-добре при сегментиране на емоционална реч от категория "гняв", отколкото моделите създадени за "гняв". Това наблюдение косвено потвърждава подобни наблюдения на публикации от други автори.

Изследванията публикувани в статията докладват разработка по международният проект PlayMancer (FP7 215839) "A European Serious Gaming 3D Environment", съфинансиран по седма рамкова програма (FP7) на Европейската Комисия.

[A5]. I. Mporas, **T. Ganchev**, O. Kocsis, N. Fakotakis: *Context-adaptive pre-processing scheme for robust speech recognition in fast-varying noise environment*, Signal Processing, vol. 91, no. 8, 2011, pp. 2101–2111. **(ISI Thomson IF 2012=1.851)**

Речева комуникация между мотоциклетист и информационна система е необходимост при някой професионални дейности (например дейности и операции на органите на реда), а също така и за съвременни забавления като достъп до Интернет, гласово управление на музикални и други лични устройства. Основните трудности за осигуряване на надеждна речева комуникация в условия на движещ се мотоциклет се дължат както на бързо променящия се шумов фон, така и на промените в речта дължащи се на вибрациите, стреса върху тялото, физическите и когнитивни усилия на водача поради повишеното натоварване. Статията е фокусирана върху разработване на метод за предварителна обработка на речевия сигнал, която да потисне максимално шумът от бързо изменящата се околна среда около движещ се мотоциклет.

- В статията се предлага метод за предварителна двустъпална обработка на речеви сигнали, който е чувствителен към контекста, и който използва динамична адаптивна селекция за да осигури максимално потискане на шума преди полезният сигнал да се подаде към последващият го модул за автоматично разпознаване на речта. Следва да се отбележи, че при предложеният метод не е необходимо, и не се извършва, разпознаване на типа шум като известна наименувана условна категория.
- Предложеният метод осъществява схема, при която динамично се избира един от множество разнородни паралелно работещи методи за потискане на смущенията, така че във всеки момент се използва най-подходящият за ситуацията алгоритъм за шумоподтискане. Изборът кой алгоритъм да се използва за конкретния момент се базира на метод за неръководено (unsupervised) групиране, посредством моделиране със смес от Гаусови функции (GMM), на дескрипторите описващи акустичното пространство и последващо определяне на най-подходящият метод за шумоподтискане чрез някаква функция на съответствията. В общия случай броят на клъстерите е различен от броя на условните категории шум. Функцията на съответствията се обучава статистически с използване на представителна за средата база данни и отчитайки полезността на всеки шумоподтискащ алгоритъм за конкретните условия. Като критерии за полезност на шумоподтискането се следи прецизността при разпознаване на думи, в модула за автоматично разпознаване на реч.
- Експериментални изследвания с базата данни "MoveOn Motorcycle Speech and Noise Database" потвърждават практическата полза от предложения метод. Наблюдава се подобрение на прецизността при разпознаване на думи с 3.3 %, в сравнение с най-добрия самостоятелен алгоритъм за шумоподтискане. Най-добри резултати са наблюдавани когато първото стъпало използва свързаните изходи на три набора от GMM с 4, 8, и 16 клъстера, и когато функцията на съответствията се изпълни с машина с опорни вектори (SVM).
- Предложеният метод не е ограничен до конкретното приложение и може да бъде полезен за други приложения, където диалогови системи се използват в бързо-изменящи се условия. Това се дължи на факта че процесът на обучение е data-driven и не се изисква предварително експертно знание за осъществяване на съответствието между определена шумова среда и конкретен метод за шумоподтискане, нито пък се изисква анотация или наименуване на условните категории шум.

Изследванията са финансирани по проекта MoveOn (IST-2005-034753) "Multi-modal and multi-sensor zero-distraction interaction interface for two wheel vehicles ON the move", съфинансиран по шеста рамкова програма (FP6) на Европейския Съюз.

[A6]. I. Mporas, O. Kocsis, **T. Ganchev**, N. Fakotakis: *Robust Speech Interaction in Motorcycle Environment*, Expert Systems With Applications, ISSN: 0957-4174, vol. 37, no. 3, March 2010. pp. 1827-1835. **(ISI Thomson IF2011=2.203)**

Управлението на превозни средства, и особено управлението на мотоциклет, са дейности които изискват пълната концентрация на водача, като в същото време водачът е със заети очи и ръце. В тази ситуация, използването на речеви интерфейс позволява на водача да управлява устройства или да ползва информационни услуги с помощта на гласови команди, без да намали значително концентрацията си и без да отмества погледа си от пътя и ръцете си от органите за управление на превозното средство. Същевременно, е необходимо да се отбележи, че работата на системите за разпознаване на реч е силно затруднена в условията на движещ се мотоциклет, с бързо изменящи се акустични условия, и повишени физически и когнитивни натоварвания на водача.

- В статията се предлага нов метод за осигуряване на надеждно автоматично разпознаване на гласови команди, в условията на бързо променяща се и силно зашумена акустична среда. Методът се основава на колаборативна схема, при която множество канали за обработка на речевия сигнал, всеки включващ модули за предварителна обработка на сигнала, шумоподтискане и разпознаване на команди, работят в паралел, а изходите им се комбинират по подходящ начин. Всеки канал за обработка на реч включва различен алгоритъм за шумоподтискане и модул за разпознаване на реч снабден с подходящо адаптиран акустичен модел. Подходящо комбиниране на изходите на отделните канали за обработка на сигнала води до компенсиране на индивидуалните несъвършенства на отделните методи за обработка на сигнала, поради което прецизността на колаборативната схема надвишава значително прецизността на всеки един от индивидуалните канали за разпознаване на реч взети поотделно.
- На базата на предложения метод е изградено цялостно решение за постигане на надеждна работа на речеви интерфейс за водачи на мотоциклети, който позволява на мотоциклетистите да ползват информационни услуги и да активират устройства (камера, радиостанция, дисплей, и други) намиращи се в шлема, в специалното компютъризирано облекло, или по мотоциклета.
- Практическата полза от предложения метод е изследвана с помощта на платформата за разработване на диалогови системи Olympus/RavenClaw, която реализира многозадачна работа и всеки компонент е изпълнен като отделен агент. Експериментите са извършени с помощта на база данни с записи на действащи служители на мотоциклетните подразделения на полицията, с реален полицейски мотоциклет и множество различни шлемове, при вдигната и свалена позиция на визьора. Конкретното приложение, за което е разработен интерфейса, е осигуряване на информационно обслужване за подразделение на мотоциклетните полицейски сили в Обединеното Кралство.
- Изследване на различни методи за комбиниране на резултатите от индивидуалните канали за обработка и разпознаване на реч, показа че предложената колаборативна схема работи най-добре когато комбинирането на изходите се извършва с алгоритъма Adaboost.M1. Измерено като точност при разпознаване на думи, подобрението е 8.0 % спрямо най-добрия индивидуален канал за обработка на реч, а измерено в брой коректно разпознати думи това предимството е 5.48 %. Предложеният подход е универсален и може да бъде използван в други подобни приложения, където се изисква надеждно разпознаване на реч в силно зашумена и бързо изменяща се акустична среда.

Изследванията са финансирани по проекта MoveOn (IST-2005-034753) "Multi-modal and multi-sensor zero-distraction interaction interface for two wheel vehicles ON the move", съфинансиран по шеста рамкова програма (FP6) на Европейския Съюз.

[A7]. I. Mporas, **T. Ganchev**: *Estimation of Unknown Speaker's Height from Speech*, International Journal of Speech Technology, ISSN 1381-2416, vol. 12, no. 4, 2009, pp.149–160.

През последното десетилетие рязко се засили интересът към определяне на биометрични индикатори, които позволяват еднозначно удостоверяване на истинската самоличност на човек. Наред с традиционните високонадеждни пръстови отпечатъци и ДНК анализ, използвани главно в съдебни разследвания, приложение намират и изменчиви биометрични показатели като цвят на кожата, цвят на очите структура на тялото, тегло, ръст, акцент и други, които сами по себе си не са уникални за дадена личност, но могат да се комбинират успешно с други биометрични индикатори за повишаване надеждността на разпознаване. Особен интерес представляват биометричните показатели, които могат да се определят от гласа на клиента или пациента (например чрез технология за разпознаване на диктори, чрез оценяване на ръста и обема на тялото на диктора, и др.), тъй като това позволява удостоверяването на самоличността, или друга личностна характеристика, да се извърши дистанционно, по телефона, без на клиента да му се натрапва някаква специална процедура или специални изисквания.

- В статията се предлага регресивен метод за директна автоматична оценка на ръста на голям брой непознати диктори от техния глас. При този метод речевия сигнал се параметризира за да се извлекат описатели на речта, които се подават към предварително създаден регресивен модел, който от своя страна директно оценява ръста на диктора. Главния фокус на изследванията докладвани в статията е определяне на приложимостта и прецизността на набор от линейни и нелинейни регресивни алгоритъма при използване на акустични дескриптори изчислени за кратки 20 ms отрязъци от входния сигнал, и глобални дескриптори определени след статистическа обработка на акустичните дескриптори. Полезността на всеки от тези дескриптори е оценена посредством алгоритъма Relief и се докладват резултати за различни подмножества на пълния набор дескриптори.
- Експериментите са проведени с използване на широко известната база данни TIMIT, която съдържа записи на 630 диктора от различни региони на САЩ. По време на експериментите е следван стандартният протокол за разделение на данните за обучение на моделите и данните за тестване на системата. Предложеният метод за директна оценка на ръста постига средна абсолютна грешка от 0.053 метра, когато регресията се изпълни с алгоритъма за Bagging регресия. Тази точност на автоматичната система отговаря на средна относителна грешка от около 3 %, докато по данни от литературата средната относителна грешка постигана от хора слушатели е около 14 %.
- Предложеният автоматичен метод е подходящ за използване в приложения свързани със автоматично наблюдение и охранителни системи, системи за дистанционни услуги по телефона и телефонни информационни центрове, където обектите на наблюдението са непознати за системата личности. Комбинацията от оценка на ръста на диктора от гласа му и оценка на височината му посредством видео-наблюдение са взаимно допълващи се, и позволява да се преодолеят някой от недостатъците на всеки от методите.

Изследванията докладвани в статията са финансирани по проекта PROMETHEUS (FP7-ICT-214901) "Prediction and Interpretation of human behaviour based on probabilistic models and heterogeneous sensors", съфинансиран по седма рамкова програма (FP7) на Европейския Съюз.

II. Глави в книги, общо 6 броя

[A8]. **T. Ganchev**, I. Mporas, N. Fakotakis: "*Audio Features Selection for Automatic Height Estimation from Speech*", S. Konstantopoulos et al. (Eds.): Advances in Artificial Intelligence: Theories, Models, and Applications, LNAI 6040/2010. Springer-Verlag Berlin Heidelberg, 2010. pp. 81–90.

Изследвания на Рентгенови и магнито-резонансни изображения (MRI) показват, че дължината на гласовия тракт и ръстът на човек са силно корелирани величини. Въз основа на това са проведени множество изследвания целящи да установят доколко точно може да се определи ръста на човек от неговия глас. За целта в научната литература са изследвани множество дескриптори на речевия сигнал, като например формантните честоти на речта, периода на основния тон, енергията на речта, кепстрални коефициенти определени посредством мел-банка филтри, коефициенти на линейно предсказване и други. Резултатите от тези изследвания показват, че макар да има директна зависимост между ръста на диктора и някой от дескрипторите на речта, нито един от тях не позволява да се определи ръста на човек с добра точност. В същото време, липсват изследвания на големи набори от дескриптори на речта и изследвания за ефективността на различни комбинации и подмножества от дескриптори на речта. Автоматичната оценка на ръста на човек по гласът му намира приложение в системи за автономен надзор и осигуряване на контролиран достъп, където удобно се комбинира с информация получена посредством наблюдение с видео и термални камери.

- В тази глава от книга се предлага обстойно изследване на относителната ефективност на голям набор от дескриптори на речта -- общо 6552 различни дескриптора и групи от дескриптори -- посредством алгоритъма известен като „регресивен Relief-F алгоритъм". Тези дескриптори на речта включват както набор от кратковременни дескриптори, представящи свойствата на кратки сегменти от сигнала (20-50 ms), така и глобални дескриптори получени за цяла фраза или изречение след статистическа обработка на кратковременните дескриптори. Подредените по важност дескриптори се групират в множества, обхващащи най-важите n дескриптора, където n ={1, 2,…, 9, 10, 20, …90, 100, 200, …, 1000}, и за всяко подмножество от n дескриптора е оценена точността при определяне ръста на диктора.
- Моделирането и оценяването на ръста на диктора по дескрипторите на речевия сигнал се извършва посредством метод за регресия, базиран на опорни вектори (SVM regression), който е обучен да изчислява директно ръста на непознати на системата диктори. Непознати, в смисъла че техният глас не е използван при създаването или оптимизацията на моделите.
- С помощта на предложеният метод за моделиране е проведено експериментално изследване на подмножествата от най-важните n (top-n) дескриптора, дефинирани по-горе. Потвърдена е важността на кепстралните коефициенти изчислени посредством мел-банка филтри, но също така е показано, че статистическите параметри производни на честотата на основния тон на речта са относително силно корелирани с ръста на диктора, докато абсолютната стойност на честотата на основния тон не е. Обобщени експерименталните резултати показват, че (i) подмножеството от top-50 дескриптора осигуряват приемлив компромис между изчисления и прецизност и най-добра прецизност, в смисъла на минимална средна абсолютна грешка и средна квадратична грешка, се наблюдава при подмножеството от top-200 дескриптора.

Изследванията са финансирани по проекта PROMETHEUS (FP7-ICT-214901) "Prediction and Interpretation of human behaviour based on probabilistic models and heterogeneous sensors", съфинансиран по седма рамкова програма (FP7) на Европейския Съюз.

[A9]. T. Kostoulas, **T. Ganchev**, N. Fakotakis: "*Affect recognition in real life scenarios*", A. Esposito et al. (Eds.): COST 2102 Int. Training School 2010, LNCS 6456, Springer-Verlag, 2011. pp. 429–435.

Снабдяването на машините с усещане за човешката емоционалност е важно за подобряване на взаимодействието човек-машина, и също подпомага автоматизираното откриване на нетипични поведения, опасност, или кризисни ситуации в системите за надзор и в приложенията за наблюдение на поведението на хора и групи хора на обществени места.

- Тази глава от книга е насочена към изследване на метод за откриване и разпознаване на определени емоционални състояния, като „паника", „гняв", „щастие" в условия близки до тези в реалния живот. Този метод се осланя на алгоритъм за параметризиране на аудио сигнала работещ на ниво на цяла фраза или изречение и схема за разпознаване на образци базиран на смес от Гаусови функции адаптирани от универсален фонов модел (UBM-GMM). Посредством избрания алгоритъм за параметризация се изчислят 384 дескриптора, като периода на основния тон, енергията на речта за текущия сегмент, кепстрални коефициенти определени посредством мел-банка филтри, отношението на хармоничните съставящи към шумовите съставящи за сегмента, техните първи производни, както и 16 статистически функции приложени върху основните параметри. Тези статистически функции (средна стойност, стандартно отклонение, ексцес, асиметрия, обхват, минимална и максимална стойност и относителна позиция, два вида регресивни коефициенти и тяхната средно квадратична грешка и др.) се изчисляват за цяла фраза или изречение. Оптималните параметри на модела бяха определени след серия експерименти с различен брой компоненти на UBM-GMM модела.
- Приложимостта на предложения метод е изследвана за сценариите представени в база данни PROMETHEUS, която е реализирана в различни затворени и открити пространства. Тези пространства и условията на експеримента симулират близки до реалните условия, като умен дом (smart-home), наблюдение на публични пространства (летища, банкомати и др.) в ситуации с един или повече участника и позволяват да се получи достоверна представа за ограниченията на метода.
- Експерименталните резултати показаха, че най-висока прецизност се постига за детектора на емоционално състояние „гняв" в условия близки до реалните при експлоатация на система за наблюдение на публични пространства. В същото време, точност на детектора (грешка от порядъка на 25 %), която макар и да е сравнима с най-добрите постижения в областта до момента, е далеч недостатъчна за практическа приложимост в реални системи, тъй като методът пропуска да открие почти една четвърт от ситуациите, в които хора наблюдатели биха открили израз на „гняв" в изказа. При емоционалните състояния "паника" и "щастие" е наблюдаван относителен дял на грешките от около 30 %. Най-вероятно това се дължи на факта че при тях интензивността на емоцията е по-малка от тази при "гняв" и следователно е по-трудно различима спрямо емоционално неутралната реч.

Понастоящем тази технология започва да намира приложение в прототипни системи и системи работещи в контролирани условия, но преди масово разгръщане на тази технология в реални приложения е необходимо значително подобрение на прецизността при разпознаване на емоционални състояния.

Изследванията са финансирани по проекта PROMETHEUS (FP7-ICT-214901) "Prediction and Interpretation of human behaviour based on probabilistic models and heterogeneous sensors", съфинансиран по седма рамкова програма (FP7) на Европейския Съюз.

[A10]. T. Kostoulas, **T. Ganchev**, N. Fakotakis: "*Study on Speaker-independent Emotion Recognition from Speech on Real-world Data*", COST2102 Workshop, Partas, Greece, Oct.29-31, 2007. Revised papers published in LNAI-5042, A. Esposito, et al. (Eds.), ISBN 978-3-540-70871-1, Springer, 2008, pp.235–242.

Множество приложения на гласовите диалогови системи, като например

- системи за автоматично пренасочване на телефонни разговори (call routers),
- гласови информационни системи на институции и бизнеса (voice portals),
- гласови информационни системи за разписания или резервации за влакове, самолети, кораби, кина, ресторанти, музеи и др.,
- системи за гласово управление в среда на умен дом (smart-home),
- системи за гласово банкиране (voice banking), и други подобни,

са предназначени да служат на големи целеви групи, с потенциално неограничен брой потребители. За постигане на добро качество на услугата и удовлетвореност на потребителя, в такива системи се изисква автоматично откриване на затруднения в комуникацията, отрицателни емоции и неудовлетвореност на клиента, след което потребителя автоматично се пренасочва към дежурен оператор-човек който да го обслужи максимално любезно. Автоматичното откриване на отрицателни емоции в такива системи поставя проблема за разпознаване на емоционални състояния за непознати за системата потребители, т.е. независимо кой е диктора.

- В тази глава се изследва поведението на детектор за отрицателни емоционални състояния в условията на симулирани емоции (актьорска интерпретация на емоционални състояния) и в условия близки до реални житейски ситуации с обикновени хора. Акцентът е върху автоматичното откриване на отрицателни емоции за непознати за системата потребители, като е предложено сравнение между три различни бази данни. Моделирането на емоционалните категории се базира на смес от Гаусови функции.
- Експерименталните резултати от тестове с различни бази данни емоционална реч показват значителни различия в прецизността на детектора между симулирани емоции и емоционални реакции в реални житейски условия. Анализирани са различията при разпознаване на симулирани емоции и емоционални реакции записани в условия близки до реалните и е обяснена разликата в прецизността на системата.
- Изследва се възможността за повишаване на прецизността на детектора посредством неравностойно комбиниране на изходите на два различни метода за класификация. Първият метод използва дескриптори на речта, изчислени за последователност от кратки сегменти от речевия сигнал, съгласно стандартните практики в системите за обработка на реч (кепстрални коефициенти, енергия на сигнала, хармоничност, честота на основния тон). При втория метод, моделите използват глобални дескриптори на речта изчислени за цяло изречение или фраза. Глобалните дескриптори са получени чрез статистическа обработка (средна стойност и стандартно отклонение) на основните кратковременни дескриптори, като към тях са добавени отношението между дължините на вокализираните и невокализираните сегменти на речта и отношението между дължините на вокализираните сегменти към дължината на цялата фраза или изречение. Отчетено е слабо подобрение на прецизността в резултат на комбинацията от двете системи.

Изследванията са финансирани по проекта PlayMancer (FP7-ICT-215839-2007) "PlayMancer: A European Serious Gaming 3D Environment", съфинансиран по седма рамкова програма (FP7) на Европейския Съюз.

[A11]. T. Kostoulas, I. Mporas, **T. Ganchev**, N. Fakotakis: „*The Effect of Emotional Speech on a Smart-Home Application*”, LNAI 5027, N.T. Nguyen et al. (Eds). © Springer-Verlag Berlin Heidelberg. pp. 305-310.

В приложения като информационни кьошкове, телефонни центрове, и системи за умен дом, където се използват гласови или мултимодални диалогови системи, основното предизвикателство е да се осигури удобен за потребителя интерфейс човек-машина, който да гарантира устойчивост и прецизност в различни условия на експлоатация и при различни поведенчески стилове на потребителите. От научната литература е известно, че емоционално оцветената реч понижава прецизността на системите за автоматично разпознаване на реч и влошава качеството на комуникация между системата и потребителя, особено в случаите когато диалоговата система не е снабдена с модул за разпознаване на емоционалната окраска на речта и липсва механизъм за смяна на стратегията за комуникация при възникване на проблем в комуникацията между машината и потребителя.

От друга страна, възможността да се разпознават най-критичните категории емоционална реч, за които грешката е най-голяма, би позволило да се вземат мерки за подобряване успеха на комуникацията. Усъвършенстване на диалоговите системи с включването на нов компонент, които да разпознава най-критичните категории емоционална реч, и да активира подходящо адаптирани модели за разпознаване на емоционална реч, които са адаптирани за най-проблематичните категории емоционална реч, би позволило повишаване на цялостната прецизност на диалоговата система.

- В тази глава от книга се изследва влиянието на различни категории емоционална реч върху прецизността на система за автоматично разпознаване на реч в приложение от типа умен дом. Целта е да се оцени за кои категории емоционална реч прецизността на системата за автоматично разпознаване на реч дегенерира най-много, спрямо типичната за емоционално-неутрална реч.
- За осъществяване на експериментите бе използвана системата за разпознаване на реч Sphinx III, която е разработена от CMU и предоставена като свободен код (open source). Използвни са стандартни акустични модели на речта създадени от емоционално неутрална реч от базата данни Wall Street Journal database. Настройките на системата за разпознаване на реч следват стандартните най-добри практики, съгласно публикации в научната литература.
- За източник на емоционална реч бе използвана базата данни на LDC "Emotional Prosody Speech and Transcripts database", която съдържа 14 категории емоционална реч (безпокойство, отегчение, хладен гняв, презрение, отчаяние, отвращение, въодушевление, радост, горещ гняв, любопитство, паника, гордост, скръб, свян) и емоционално неутрална реч. Речевия сигнал за всички категории изброени по-горе бе обработван и параметризиран по еднакъв начин. Настройките на системата за разпознаване на реч и експерименталния протокол бяха еднакви за всички категории.
- Експерименталните резултати показват че прецизността на разпознаване на думите се различава в голяма степен за различните категории емоционална реч. Най-ниски стойности на грешката при разпознаване на думи е при емоционално неутрална реч, 6.2%, а най-високи при емоционална категория "горещ гняв", 44.1 %, следвана от "въодушевление" с 36,3 % и "паника" с 35.2 %. Категориите реч с по-слаба интензивност на емоцията показват по-ниски нива на грешката. Това знание създава условия за планиране на бъдещи подобрения, чрез създаване на специализирани акустични модели адаптирани за критичните категории "горещ гняв", "въодушевление" и "паника".

Изследванията са финансирани по проекта LOGOS (ЕНГ-102) "A general architecture for speech recognition and user friendly dialogue interaction for advanced commercial applications", финансиран от Генералният секретариат за наука и технологии на Гърция.

[A12]. I. Mporas, **T. Ganchev**, P. Zervas, N. Fakotakis: *Recognition of Greek Phonemes using Support Vector Machines*, G. Antoniou et al. (Eds.): SETN-2006, LNAI 3955, © Springer-Verlag Berlin Heidelberg 2006. pp. 290-300.

Съвременните многоезични гласови интерфейси задължително включват модул за автоматично разпознаване на езика по фраза или изречение, което позволява автоматизация и удобства за потребителя. В съвременните системи за разпознаване на реч и език фонотактичният метод намира широко приложение тъй като предлага добър компромис между прецизност при разпознаване и обем на базата данни необходима за създаване на модела. При фонотактичния метод, речевият сигнал се декодира до поредица от фонеми, която се подава към модел на езика (в англоезичната научна литература методът е известен като phone recognizer followed by language model, PRLM). Заради недостатъчната прецизност на фонемните модели обаче при реализацията на метода за всяка инстанция от времето не се използва само една фонема, изчислена като най-вероятната, а към модела на езика се подава списък от най-вероятните n фонеми (n-best).

- В тази глава се изследва приложимостта на класификатор базиран на опорни вектори (SVM) за разпознаване на фонеми. Фокусът е върху LS-SVM вариант на класификатора, при който моделът се обучава по метода на най-малките квадрати, сравнително икономично и надеждно, в сравнение с други варианти на SVM. Изследването е ограничено до създаването на фонемен модел, поради което модел на езика не се разглежда.
- В експериментите е използвана база данни SpeechDat (II) FDB-5000-Greek, като първоначално се изследва приложимостта на два набора от дескриптори на речта. Първият набор е съставен от 13 кепстарлни коефициента (MFCC) изчислени за кратки сегменти от речевия сигнал посредством мел-банка филтри. Втория набор включва също и 3 формантни честоти, нормализирани спрямо първата форманта F1, изчислени за всеки сегмент от сигнала.
- LS-SVM моделът е сравнен при равни условия с шест други модела, обучени със същите данни: наивен класификатор на Бейс (Naïve Bayes), невронна мрежа многослоен перцептон (MLP NN), регресивно дърво на решенията от типа M5P, дърво на решенията от тип C4.5, линеен класификатор с опорни вектори (стандартен SVM). Тези методи са използвани самостоятелно или комбинирани посредством алгоритми за мета-класификация като Bagging и AdaBoost.
- Двата модела за класификация посредством опорни вектори показаха най-висока прецизност. Наборът дескриптори съставен от MFCC и 3 формантни честоти, нормализирани спрямо F1, показа по-добри резултати. При използване на набора дескриптори само от MFCC и отчитане само на най-вероятната фонема за всяка инстанция, прецизността е едва 34.8 %. С добавяне на 3-те форманти прецизността се покачва до 38.1%. За случая когато отчитаме първите 5-best фонеми, LS-SVM модела дава най-висока прецизност, 74.2 %, без да се използва модел на езика.

Добавянето на модел на езика, създаден за нуждите на конкретно приложение, би повишило допълнително прецизността на разпознаване.

[A13]. I. Mporas, **T. Ganchev**, N. Fakotakis: "*Phonotactic Recognition of Greek and Cypriot Dialects from Telephone Speech*", J. Darzentas et al. (Eds.), LNCS-5138, ISBN: 978-3-540-87880-3, Springer Berlin Heidelberg, 2008, pp.173–181.

Разпознаването на диалекти на езика е изключително важно за постигане на висока прецизност при разпознаването на реч в автоматизираните диалогови системи с гласов или мултимодален интерфейс. Правилното разпознаване на диалекта позволява на диалоговата система да превключи към подходящо адаптирани акустични модели и граматични правила, за да се избегнат недоразумения при общуването човек-машина. Разпознаването на диалект е по-трудна задача в сравнение с разпознаването на език, но поради сходството им често методи създадени за автоматично разпознаване на езика се адаптират за разпознаване на диалекти. Разпознаването на диалект използва няколко различни нива на описание на езика: акустично (спектрална информация), прозодично (прозодия), фонотактично (модели на езика), и лексикална информация. Статистическите методи за разпознаване на диалекти разчитат на големи бази данни, но поради големите разходи, за част от Европейските езици такива ресурси не са налични или са недостатъчни, което затруднява разработването на гласови интерфейси.

- В тази глава се оценяват различни конфигурации на фонотактична система за разпознаване на диалект и се изучава приложимостта на комбиниран метод базиран на архитектура с паралелни фонемни модели последвани от модел на езика (PPRLM). При този метод се използват набор от различни акустични модели за разпознаване на фонеми, създадени за различни Европейски езици (чешки, руски, унгарски, гръцки, английски и немски), и модели на езика за модерен гръцки диалект и за кипърски диалект. Акустичните модели са базирани на скрити модели на Марков (HMM) и са създадени от големите бази данни SpeechDat-E и SpeechDat(II). Моделите на езика са n-gram модели създадени от фонетичните транскрипции на съответните бази данни с помощта на CMU Cambridge Statistical Language Modeling (SLM) Toolkit.
- За изследване на приложимостта на различните конфигурации и методи за разпознаване на диалект бяха използвани базите данни SpeechDat(II) FDB-5000-Greek и Orientel Cypriot Greek database със записи на телефонни разговори на приблизително 5000 и 1000 диктора. Следва да се отбележи, че PPRLM метода показа значително по-висока прецизност в сравнение със стандартният подход, при който се използва само гръцки акустичен модел. Една от причините е в недостатъчният ресурс от подходящи бази данни за гръцки език, които да позволят висококачествени акустични модели за разпознаване на фонеми, което отчасти се компенсира от свръх висококачествените модели за немски и английски, които са по-чувствителни към малки вариации в произношението на определени фонеми. Поради това, най-висока точност на разпознаване между модерен гръцки и кипърски диалекти се получават когато се използва PPRLM метода с комбиниране на резултатите от английски, немски и гръцки акустични модели. Добавянето на акустични модели за чешки, руски и унгарски не повишава прецизността на системата, главно поради значителните разлики във фонетичната структура спрямо гръцки език.
- Експерименталните резултати за разпознаване на двата основни диалекта на гръцки език: модерен гръцки език и кипърски гръцки показват високата прецизност на PPRLM метода, която достига до 95 % и е достатъчна за внедряване в практически приложения, като гласово банкиране, гласови портали, информационни системи, мобилни телефонни услуги и др.

III. Рецензирани доклади от конференции в сборник доклади (7 броя)

[A14]. I. Mporas, **T. Ganchev**, O. Kocsis, N. Fakotakis: "*Dynamic Selection of a Speech Enhancement Method for Robust Speech Recognition in Moving Motorcycle Environment*", Proc. of the ICASSP-2011, Prague, Check Republic, pp. 5176-5179.

В системи за гласов контрол на услуги или устройства, като например системата MoveOn, се изисква ползването на информационни услуги по време на предвижване с мотоциклети. Прецизността на системите за разпознаване на реч се влошава значително в отворени пространства, на улицата, и особено при предвижване с мотоциклет, където акустичната среда е силно зашумена и бързо променяща се. В такива условия, към полезния речеви сигнал са насложени множество адитивни нестационарни смущения, които се търпят значителни изменения във времевата и спектралната област. Тъй като в приложения на гласов контрол командите и фразите са кратки (до 3 секунди) може да се приеме с известно приближение, че в рамките на една фраза акустичните условия не се изменят значително. Между последователни фрази обаче е възможно значително изменение на обстановката.

- В доклада е разгледан метод за предварителна обработка на речевия сигнал с цел потискане на адитивни смущения и постигане на добра прецизност при последващото разпознаване на гласови команди. Предлага се метод за шумопотискане с динамичен подбор на най-подходящия алгоритъм в зависимост от акустичната среда, с цел да се подобри прецизността при разпознаването на реч. За целта акустичната среда се моделира със смес от Гаусови функции (GMM), като моделът е съставен от база данни представителна за условията в които ще работи разпознаването на реч. Моделът се избира да е от нисък ред и разделя акустичното пространство на няколко области, които не е задължително да съответстват на някакви наименувани категории. По време на работа, порции от входния акустичен сигнал се сравнява с модела и според степента на съвпадение се избира един от наличните методи за шумоподтискане. Конкретно изборът се осъществява посредством функция на съответствието, която се обучава да избира определен алгоритъм в зависимост от изхода от сравнението на входния сигнал и GMM модела. За целта предварително експериментално е установена приложимостта на всеки метод за набор от типични условия на акустична среда в която ще работи системата.
- При експерименталната оценка на предложения метод са подбрани 8 алгоритъма за шумоподтискане (спектрално изваждане, спектрално изваждане с оценка на шума, много-лентово спектрално изваждане, минимален средно квадратичен логаритмично-спектрален оценител на амплитудата, Бейс оценител, и други) и четири различни реализации на функцията на съответствието: класификация чрез опорни вектори (SMO), невронна мрежа многослоен перцептон (MLP NN), дърво на решенията от тип C4.5 (J48), и класификатор по k най-близки съседи (IBk).
- Експерименталните резултати получени посредством базата данни MoveOn показват, че предложеният метод за динамичен избор на шумоподтискащ алгоритъм за всяка фраза/команда, подобрява прецизността при разпознаване на думи с 3.3 %, спрямо най-успешният измежду всички методи за шумоподтискане работещ самостоятелно. За специфичните условия на базата данни MoveOn, най-подходящият размер на GMM модела е 28, а класификацията чрез опорни вектори (SMO) е най-успешната реализация на функцията на съответствието за избор на метод за шумоподтискане в зависимост от смущенията присъстващи във входния сигнал.

[A15]. **T. Ganchev**, I. Mporas, N. Fakotakis: "*Automatic Height Estimation from Speech in Real-World Setup*", Proc. of the 2010 European Signal Processing Conference (EUSIPCO 2010), Aalborg, Danmark, August 23-27, 2010. pp. 800-804.

Методите за оценка ръста на диктор от произнесена фраза или изречение се основават на допускането, че има строга зависимост между ръста на човек и дължината на неговия гласов тракт. Според модела на Фант и съвременната теория за производство на реч, дължината на гласовия тракт се корелира добре със стойностите на формантните честоти, което позволява да се определи ръста на човек от дескрипторите на речта. Експериментално е доказано, че само четвъртата формантна честота е корелирана с ръста на човек, но други дескриптори на речта, като коефициенти на линейно предсказване, кепстрални коефициенти, височина на основния тон на речта, енергията, и др. също са повече или по-малко корелирани с ръста на диктора. Практическите приложения на тази технология е в автономни системи за наблюдение и надзор на публични пространства и системи от типа умен дом.

- В доклада се издига хипотезата, че тъй като не е известен един определен дескриптор на речта, който да осигури прецизна оценка на ръста на човек по изказана фраза или изречение, автоматичните системи за оценка на ръста ще трябва да разчитат на голям набор от корелирани с ръста и желателно комплементарни един спрямо друг дескриптори. Очаква се комбинацията от внимателно подбрани дескриптори да доведе до висока прецизност на оценката на ръста.
- Предложен е метод за директна оценка на ръста на непознати диктори базиран на регресивен анализ с Гаусов процес. Предложеният метод използва дескриптори на речта определени за цяла фраза или изречение, които се използват за създаване на регресивен модел за оценяване на ръста на диктора и интервала на несигурност на тази оценка.
- Експерименталното изследване е проведено на базите данни TIMIT и PROMETHEUS. TIMIT е със студийно качество и е най-често използваната база данни защото осигурява стандартен експериментален протокол за сравнение с други методи. PROMETHEUS позволява оценка на работоспособността на метода в множество сценарии в помещения, публични пространства и на открито, съответстващи на реални експлоатационни условия за приложения като автономни системи за наблюдение и надзор на публични пространства като летище, банкомат и др., а също и системи от типа умен дом.
- Експерименталните резултати показват, че методът за директна оценка на ръста на непознати диктори, базиран на регресивен анализ с Гаусов процес с нормализирано полиномиално ядро, показва най-устойчива и прецизна работа. Експериментите за изчисляване степента на корелираност на всеки дескриптор на речта с ръста, проведени с базата данни TIMIT и последващите експерименти за оценяване на ръста при нарастващ брой дескриптори (top-n) за n=10,20,30,...,100, 200, ...1000, показват че първите 50 (top-50) дескриптора осигуряват добър компромис между изчисления и точност, а най-висока точност се получава при наборът от дескриптори съставен от първите 300 (top-300) дескриптора. Добавянето на повече дескриптори не подобрява прецизността, тъй като те не носят допълнителна информация извън това което е описано от top-300. Получените резултати подкрепят издигнатата хипотеза.
- Усреднената относителна грешка при оценяване на ръста е около 3 % за двете бази данни и за помещения и отворени пространства, което е добро потвърждение за устойчивостта на метода. Обаче следва да се отбележи, че с понижаване на отношението сигнал шум под 6 dB неточността нараства прогресивно, като най-силен ефект имат адитивни смущения от конкурентен диктор.

Изследванията са финансирани по проекта PROMETHEUS (FP7-ICT-214901) "Prediction and Interpretation of human behaviour based on probabilistic models and heterogeneous sensors", съфинансиран по седма рамкова програма (FP7) на Европейския Съюз.

[A16]. I. Mporas, **T. Ganchev**, M. Siafarikas, T. Kostoulas: „*Comparative Evaluation of Speech Parameterizations for Speech Recognition*”, Proc. of the ICTAI-2007, October 29-31, 2007, Patras, Greece, pp.510-513.

Съвременните методи за разпознаване на реч са базирани на статистически анализ на речта, посредством могъщи техники за разпознаване, като скрити Марковски модели (HMM) и динамично програмиране, и скрити Марковски модели (HMM) комбинирани с невронни мрежи или SVM, динамични Бейсови мрежи и др. Един проблем който остава неразрешен до момента е параметризацията на реч. Параметризацията на речта цели извличането на дескриптори които да представят полезната информация в компактна форма, така че да може ефективно и прецизно да се открие съответствието между модел и входен сигнал.

- В доклада се извършва сравнително изследване на практическата полза от някой скорошни разработки на нови дескриптори на речта и приложимостта им при разпознаване на реч.
- Използван е общ експериментален протокол за да се сравнят прецизността при разпознаване на реч, получена за някой по-слабо изследвани дескриптори на речта, базирани на дискретното преобразуване с уейвлет пакети (DWPT), и традиционни дескриптори като кепстрални коефициенти определени посредством мел-банка филтри (MFCC) и коефициенти на линейно предсказване (PLP), които понастоящем са най-широко използвани. Изследвани са общо единадесет различни вида дескриптори, измежду които уейвлет дескрипторите на Sarikaya и Hansen за подлентово кодиране (SBC), дескрипторите с уейвлет пакети на Farooq и Datta (WPF), дескрипторите с уейвлет пакети за разпознаване на диктори (WPSR) в две разновидности, а също така няколко различни модификации на капстрални коефициенти и накрая PLP.
- За осъществяване на експериментите бе използвана базата данни TIMIT и системата за разпознаване на реч Sphinx III, разработена от Carnegie Mellon University и предоставена като свободен код (open source) в публичното пространство. За всеки набор от дескриптори на речта са създадени акустични модели използвайки тренировъчната част на TIMIT. Настройките при обучението на акустичните модели и при тестването на системата за разпознаване на реч следват стандартните най-добри практики, съгласно публикации в научната литература. Създадени са непрекъснати акустични модели (в смисъл на неквантувани по стойност дескриптори), които използват стандартен набор от 38 фонеми и пауза, моделирани с контекстно независими HMM модели с три състояния и еднопосочна архитектура (left-to-right), като за всяко състояние са използвани 8 компонентна смес от Гаусови функции.
- Експерименталните резултати показват, че всички изследвани дескриптори превъзхождат стандартизираните MFCC. Най-ниска стойност на грешката при разпознаване на думи е отчетена за SBC, последвана от другите дескриптори базирани на преобразуване с уейвлет пакети (WPSR125, WPSR250 и WPF), и след това се подреждат всички дескриптори базирани на дискретното преобразуване на Фурие (DFT). Превъзходството на уейвлет базираните дескриптори се дължи на по-добре балансираната им локализираща способност по време и честота, по подходящия набор от базисни функции, които са добре пригодени за обработка на нестационарните речеви сигнали, в сравнение със sin(.) и cos(.) използвани в DFT.

Изследванията са финансирани по проекта MoveOn (IST-2005-034753) "Multi-modal and multi-sensor zero-distraction interaction interface for two wheel vehicles ON the move", съфинансиран по шеста рамкова програма (FP6) на Европейския Съюз.

[A17]. M. Siafarikas, **T. Ganchev**, N. Fakotakis: "*Wavelet Packet Bases for Speaker Recognition*", Proc. of the ICTAI-2007, October 29-31, 2007, Patras, Greece, pp.514-517.

Традиционните дескриптори на речевия сигнал, измежду които стандартизираните кепстрални коефициенти с мел- банка филтри (MFCC) и някой дескриптори основаващи се на дискретното преобразуване с уейвлет пакети, апроксимират нелинейното възприятие на честотата характерно за човешкото слухово възприятие. В специализираната научна литература е добре документирано, че този подход осигурява определени удобства и води до дескриптори на речевия сигнал, които превъзхождат по-ранните дескриптори основаващи се на линейна скала на честотата. Въпреки, че дескрипторите изчислени по методите следващи този подход са добре мотивирани от характеристиките на човешкия слухов апарат, слабо е изучено дали те са оптимални за нуждите на разпознаване на диктори. Проблемът е в това че човешкия слухов апарат е с общо предназначение и формирането му е компромис от необходимостта да подпомага разнородни задачи, като: "кой говори?", "какво казва?", "с каква емоционална окраска?", "на какъв език?", "с какъв диалект?" и др. Различните задачи обаче, имат взаимно-противоречащи си изисквания, като например докато разпознаването на реч се стреми да потисне информацията за диктора, т.е. за "кой говори?", за да може да се разпознава безпогрешно лингвистичната информация, то разпознаването на диктори изисква да се потисне лингвистичната информация за да се открие кой говори, независимо от това какво казва, и т.н.

- В доклада се предлага метод за обективно разделение на честотното пространство, така че да се получат дескриптори на речта, които подчертават уникалността на всеки отделен глас. Основната разлика спрямо предишни подходи, които оценяват използването на уейвлет коефициентите в дадено разклонение от дървото спрямо родителския възел от който произлизат, е че при предложеният подход се оценява цялостното разделение на честотното пространство, т.е. цялостното дърво, включващо разклоненията на дървото в цялата честотна лента. Това позволява да се получи реална оценка доколко дадено разклонение носи допълваща (комплиментарна) информация към вече наличната.
- Предложен е алгоритъм за разделяне на честотното пространство посредством дискретното преобразуване с уейвлет пакети (WP) и дървовидна структура (WP дърво), оптимизирана за разпознаване на диктори. Оптимизацията се извършва с итеративна процедура, при която последователно се оценяват всички възможни разделения на честотното пространство, като основният критерии за оценка адекватността на дадено разделение е грешката при автоматично разпознаване на диктори, получена за конкретното разделение. Такова разделение на честотното пространство спомага за определянето на дескриптори на речта, които са подходящи за разпознаване на диктори.
- Практическата стойност на предложения метод е оценена експериментално с помощта на специализираната база данни за разпознаване на диктори Polycost и разработената от автора системата за верификация на диктори, WCL-1. Полученото по предложения метод разделение на честотното пространство съдържа много повече честотни ленти (80), спрямо традиционните методи. Дескрипторите на речта определени посредством това разделение превъзхождат по прецизност както стандартизираните MFCC, така и други четири вида дескриптори използващи дискретното преобразуване с уейвлет пакети, но апроксимиращи честотно разделение с WP дърво мотивирано от свойствата на човешкото слухово възприятие.
- Предложения подход за разделение на честотната лента посредствен обективен критерии може да бъде прилаган за други задачи за класификация, където е необходимо да се оптимизират дескрипторите на речевия сигнал.

[A18]. **T. Ganchev**, N. Fakotakis, G. Kokkinakis: "*The WCL-1 System in the 2003 NIST Speaker Recognition Evaluation and 2003 NFI/TNO Forensic Speaker Recognition Evaluation*", Proc. of the SPECOM-2006, 25-29 June 2006, pp.109-114.

През изминалите две десетилетия научната общност занимаваща се с разработка на методи и технологии за разпознаване на диктори редовно провежда кампании за оценяване на напредъка постигнат от различните колективи работещи по проблематиката. В тези кампании се използват специално разработени за целта експериментален протокол и база данни. Кампаниите организирани от Националният институт за стандарти и технологии на САЩ са фокусирани на разпознаване на диктори от реч записана през наземни кабелни мрежи, клетъчни мрежи, или разнородни източници. Кампанията организирана от Холандският Национален съдебен институт и Института по технологии за човека е фокусирана върху технологии за разпознаване на диктори за нуждите на право-охранителните органи и съдебните разследвания. Обмена на информация и идеи по време на тези кампании съдействат за напредъка на технологиите за разпознаване на диктори.

- В доклада се представят резултати от участието на разработената от автора система, WCL-1, в две различни кампании за оценка на технологиите за разпознаване на диктори: 2003 NFI/TNO Forensic Speaker Recognition Evaluation (SRE) и 2003 NIST Speaker Recognition Evaluation. Тези две кампании отстоят във времето в рамките на шест месеца, поради което системата WCL-1 участва в двете без изменения. Двете кампании са от различна сложност и имат различен фокус. Анализ на резултатите получени от тези различни по цели и организация кампании позволява да се сравни работата на системата в различни условия, за нуждите на различни приложения.
- В 2003 NFI/TNO Forensic SRE бе използвана специална база данни реч състояща се от записи на реални заподозрени, направени без тяхното знание, със специализирана апаратура за подслушване на телефонни разговори от GSM до GSM. Записите са направени за нуждите на реални полицейски разследвания от специализираните служби в Холандия в продължение на две години. Колективите участниците в кампанията за разпознаване на диктори нямаха достъп до лични данни на заподозрените и целта на кампанията беше да се открие дали в два или повече записа дикторът е един и същ.
- 2003 NIST SRE включваше 5 различни задачи, като разработената от автора система WCL-1, участва само в една от тях: "One-speaker detection", която се състои във верификация на диктори. Целевата група се състои от около 350 потребителя (мъже и жени), които се моделират от системата. Опити за получаване на достъп се извършват както от потребители на системата така и от множество неизвестни за системата гласове, включително и от натрапници. Базата данни използвана по време на кампанията е компилация от извадки от голямата база данни Switchboard-Cellular Corpus Part 2 и Part 3, записани през различни телефонни мрежи в САЩ със съгласието на потребителите. Записите се отличават с голямо разнообразие на използваните телефонни мрежи и устройства и разнородни условия на околната среда.
- Както и при другите системи участвали в тези кампании, системата WCL-1 показа по-висока надеждност в ситуации, където диктора винаги използва личния си телефон, и когато говори на един и същ език. Детайлен анализ е наличен в източници [3][10][11] посочени в доклада. В тези източници, WCL-1 е система "15" в [10], и "T2" в [3] (модификация на WCL-1 е означена като "S18" в [3]).

По време на кампанията 2003 NFI/TNO Forensic SRE, създадената от автора система WCL-1 се класира втора измежду 15 участници от университети и организации от цял свят. Основното и предимство на системата WCL-1 е, че прецизността на верификацията не се влияе значително от дължината на произнесената фраза.

[A19]. **T. Ganchev**, M. Siafarikas, N. Fakotakis: "*Evaluation of speech parameterization methods for speaker recognition*", Proc. of the Acoustics-2006, Sept 18-19, 2006. pp.105-110.

Последното десетилетие отбеляза появата на множество нови методи за параметризация на речеви сигнали, базирани на дискретното преобразуване с уейвлет пакети, за които е докладвано, че имат предимство пред стандартизираните дескриптори на речта, известни като кепстрални коефициенти с мел-банка филтри (MFCC). Публикуваните изследвания обаче използват разнородни бази данни и различни експериментални протоколи, поради което трудно може да се съди за относителната полезност на предложените дескриптори. Това затруднява изборът на подходящ набор дескриптори за всеки отделен случай, поради което разработчиците на системи за обработка на реч най-често предпочитат да използват добре проучените и познати MFCC.

- В доклада се представят резултати от експериментално изследване на пет вида дескриптори на речта, както и известен брой техни подмножества, което цели да определи приложимостта на тези дескриптори за нуждите на приложения изискващи разпознаване на диктори. За по-добро разбиране на експерименталните резултати е направено сравнение със стандартизираните MFCC, които са най-често използваните дескриптори на речта. Това сравнително изследване има за цел да установи най-подходящите дескриптори на речта за приложения, чиито организационен план съвпада с този на използвания експериментален протокол.
- В експериментите с различните дескриптори на речта е използван стандартен експериментален протокол, дефиниран от Националния институт за стандартизация и технологии на САЩ за нуждите на кампанията 2001 NIST Speaker Recognition Evaluation, както и специалната база данни предоставена от NIST за нуждите на кампанията. Всички видове дескриптори на речта са изследвани при еднакви условия с използването на системата за верификация на диктори WCL-1. Сравнението между отделните видове дескриптори се осъществява по два критерия: балансирана грешка при верификация (EER) и оптимална цена за решението (DCFopt).
- Изследвани са оригиналните дескриптори, така както са дефинирани от авторите, техни под-множества, а също така и нормализирани варианти с нулева средна стойност и унифицирана дисперсия на всички кепстрални коефициенти.
- Експерименталните резултати показват, че при дескрипторите на речта изчислени посредством дискретното преобразуване с уейвлет пакети, изключването на първите три кепстрални коефициента води до значително намаляване на грешката. Това се обяснява с наблюдението, че първите коефициенти са по-чувствителни към вариациите дължащи се на канала за връзка и лингвистичната информация в речта. При MFCC изключването на първия коефициент (с индекс 0) има същия ефект.
- Експериментите показват че нормализирането на динамичния диапазон на оригиналните дескриптори води до намаляване на грешката при верификация.
- Все пак, най-добри резултати се получават при изключване на първите три коефициента, като нормализацията на това подмножество от дескриптори увеличава грешката, защото премахва важна за разграничаването на гласовете информация.
- Наборът дескриптори известен като Overlapping wavelet packet features [10], показа най-ниска грешка (EER) и конкурентни стойности за DCFopt, поради което е най-подходящия за разпознаване на диктори (измежду шестте вида дескриптори изследвани в доклада). Тези дескриптори са получени от неортогонално преобразуване с уейвлет пакети, но това не пречи в приложения където речевият сигнал няма да бъде възстановяван.

[A20]. **T. Ganchev**, P. Zervas, N. Fakotakis, G. Kokkinakis: "*Benchmarking Feature Selection Techniques on the Speaker Verification Task*", Proc. of the CSNDSP-06, 19-21 July, 2006. pp. 314-318.

Верификацията на диктори е класически пример за направлявано обучение на модели, при което моделите се създават от предварително подготвен масив от дескриптори на речта. Използването на много-измерни масиви от дескриптори повишава изискванията към обема данни за обучение на устойчиви модели, и повишава изискванията към изчислителната мощ за обучение на моделите и изискванията за налична оперативна памет. Поради тези неудобства при създаване на системи за верификация на диктори е важно да се подбере подмножество от дескриптори, които са уместни и допринасят за повишаване на прецизността на верификацията, като в същото време се избягват дескриптори които не са уместни, имат твърде голяма изменчивост в рамките на дадена категория, или които се не носят допълнителна информация, спрямо вече избраните. За осъществяването на тази задача се използват методи за оценяване и сортиране по важност на индивидуалните параметри на вектора от дескриптори. Това дава възможност да се направи компромис между прецизност на верификацията и изчислителни загуби.

- В доклада се изследва се практическата полезност на пет алгоритъма за оценяване и сортиране на дескрипторите на речта с цел подобряване на прецизността на верификация на диктори. Крайната цел е да се подбере набор от дескриптори на речта, които се състои само от уместни и комплементарни един към друг параметри на речта. Изключването на излишните дескриптори, които не допринасят за подобряване на прецизността, намалява изчислителните загуби и подобрява използването на наличния набор от тренировъчни данни. Това от своя страна допринася за подобряване на прецизността и устойчивостта на системата за разпознаване на диктори.
- Конкретното изследване е направено като част от подготовката на системата WCL-1 за кампанията 2004 NIST Speaker Recognition Evaluation, организирана от Националния институт за стандартизация и технологии на САЩ. В експериментите с петте различни алгоритъма за определяне уместността на дескрипторите на речта е използван стандартен експериментален протокол, дефиниран от Националния институт за стандартизация и технологии на САЩ за нуждите на кампанията 2001 NIST Speaker Recognition Evaluation, както и специалната база данни предоставена от NIST за нуждите на кампанията.
- След сортиране на дескрипторите на речта според приноса им за подобряване на прецизността при верификация, е извършено цялостно изследване на потенциала за подобряване на бързодействието и прецизността на системата, при избор на подмножества от дескриптори с различна размерност.
- Експерименталните резултати показват, че в голямата си част различните алгоритми за определяне на важността на дескрипторите дават сходни резултати -- с най-голям принос са логаритъма от енергията на речта, MFCC(1) и логаритъма от честотата на основния тон. С малки изключения, индекса на MFCC дава груба представа за важността на съответния дескриптор.
- При избора на размерност, най-добри резултати се получават при използване на top-28 дескриптора. Логична алтернатива дава използването на top-25 дескриптора. Като цяло обаче подобрението е малко спрямо използването на пълния набор от 33 дескриптора (MFCC с индекси от 1 до 31, логаритъм от енергията на сегмента реч и логаритъма на честотата на основния тон).

Б. Публикации извън групата равностойни на монографичен труд

I. Статии в научни списания, общо 6 броя

[Б1]. **T. Ganchev**: *Enhanced Training for the Locally Recurrent Probabilistic Neural Networks*, International Journal of Artificial Intelligence Tools, ISSN: 0218-2130, vol. 18, no. 6, 2009. pp. 853–881. **(ISI Thomson IF 2012=0.250)**

Локално рекурентните невронни мрежи имат ценната способност да моделират времеви зависимости във входните данни, което позволява да се реализира контекстно разглеждане на входната информация за всеки един момент от времето. Това им дава значително предимство пред праволинейните системи, в които информацията тече еднопосочно от входа към изхода, и в практически задачи свързани с обработката на сигнали представляващи последователности от събития. Примери за такива приложения са: идентификация на нелинейни системи, системи за управление на промишлени обекти, предсказване на времето, защита на енергийни системи, предсказване на силата и посоката на вятъра, граматична дедукция, разпознаване на реч, верификация на диктори, синтез на реч, потискане на смущения и др. Използвайки локално рекурентните вероятностни невронни мрежи (LRPNN) в различни практически приложения бяха забелязани някой недостатъци при обучението, свързани със скоростта на сходимост и използваната стратегия за постигане на баланс между класовете, главно в приложения където трябва да се балансира обучението за множество категории, които се припокриват в многомерното пространство на дескрипторите.

- В статията се предлага нова стратегия за обучение на рекурентния слой на LRPNN. Тази стратегия за обучение вече е ориентирана към оптимизиране на постериорната вероятност изчислена за последователност от входни данни за всички целеви категории. При тази стратегия, вече не се налагат наказания за несиметричност при обучение на моделите за различните целеви категории, а директно се цели увеличаване на постериорната вероятност за целевата категория и минимизиране на постериорните вероятности за останалите категории. Новата стратегия за обучение изисква по-малко усилия за изчисляване на целевата функция, което води до значително ускоряване на обучението.
- Практическата полезност на новата стратегия е оценена с използването на различни методи за определяне на теглата на коефициентите в рекурентния слой на LRPNN, като най-бърза сходимост на обучението бе постигнато за обучение базирано на оптимизация с рояк от частици (Particle Swarm Optimization, PSO), агенти които групово претърсват пространството на решенията. Новият метод води до намаляване на изчисленията, устойчивост на обучението и подобрена прецизност на LRPNN в различни конфигурации и приложни задачи: разпознаване на емоционалното състояние на диктора, разпознаване на диктори и верификация на диктори. Новата стратегия и новия метод за обучение на LRPNN се оказаха особено успешни при разпознаване на диктори, като относителната грешка на системата за разпознаване на диктори намаля с 30%. На задачата за разпознаване на емоционални новия метод за обучение изисква само около 30% от изчисленията необходими за предишния мето, като същевременно се постига понижаване на грешката и подобряване на стабилността на обучението. При задачата за верификация на диктори намалението на броя на изчисленията е приблизително 50% при запазване на прецизността на системата.

[Б2]. R. Saeidi, H.R. Sadegh Mohammadi, **T. Ganchev**, R.D. Rodman: *Particle Swarm Optimization for Sorted Adapted Gaussian Mixture Models*, IEEE Transactions on Audio Speech and Language Processing, IEEE, USA, ISSN: 1558-7916, vol. 17, no. 2, 2009. pp. 344–353. **(ISI Thomson IF 2012=1.675)**

Съвременните системи за разпознаване на диктори са базирани на статистическо моделиране на речта, което най-често се изразява в изграждане на универсален фонов модел (UBM), представляващ смес от Гаусови функции (GMM), от който в последствие се получават моделите на отделните диктори посредством адаптация на параметрите. Тази процедура, известна като UBM-GMM, води до естествена асоциация между UBM и адаптираните GMM за отделните диктори, заради това че за всеки компонент на UBM съществува съответния адаптиран компонент на специфичния за даден диктор GMM. По време на верификацията, се изчислява близостта между входния сигнал и всеки от двата модела, които обикновено са с хиляди компоненти (1024, 2048, 4096), поради което времето за изчисляване е значително. Използвайки гореспоменатото съответствие между UBM и GMM, с цел намаляване на броя на изчисленията, авторският колектив въведе модификация на метода, наречена сортирана смес от Гаусови функции (SGMM), при който може да се балансира между прецизност и брой изчисления. Известно е, че прецизността на SGMM зависи от подходящият избор на сортиращата функция и правилното определяне на параметрите и.

- В статията се предлага метод за определяне на параметрите на сортиращата функция, който се основава на цикъл на оптимизация на параметрите и с помощта на рояк от частици (Particle Swarm Optimization, PSO) и на подходяща целева функция. Тази оптимизация цели да се настрои достоверно представянето на относителната важност на отделните дескриптори на речта, с цел да се подобри адекватността на сортиращата функция и оттам цялостната прецизност на системата за верификация на диктори. Успехът на оптимизацията зависи от представителността на наличния набор от тренировъчни данни. С предложеният метод се очаква, че посредством максимизиране на целевата функция чрез PSO, можем да оптимизираме стойностите на тегловните коефициенти на сортиращата функция по такъв начин, че съседните в многомерното пространство вектори от дескриптори на речта, да бъдат представени като близки съседи, или почти съседи, по стойностите на сортиращата функция. Последното ще доведе до еднозначност и точност при намирането на регионите в които да се претърсва за коректните смеси от Гаусови функции и до намаляване на броя на изчисленията необходими за тестване на даден входен сигнал без да се понижи значително прецизността при верификация на диктори.
- Практическата полза от предложеният метод е проверена с помощта на базата данни 2002 NIST Speaker Recognition Evaluation database (SRE) следвайки стандартния експериментален протокол дефиниран от NIST.
- Експерименталните резултати показват предимството на предложения метод за определяне на параметрите на SGMM чрез PSO спрямо оригиналния SGMM. От тези резултати може да се заключи, че PSO оптимизацията намалява разликата в прецизността между оригиналния UBM-GMM и предложения метод, т.е. в сравнение с оригиналния SGMM, предложения тук SGMM, оптимизиран с PSO, компенсира до известна степен загубата на прецизност дължаща се на ускоряването на изчисленията. Двукратно се ускорява изчислението спрямо стандартния SGMM, при подобряване на прецизността.
- По нататъшно сближаване на прецизността до тази на традиционния UBM-GMM може да се получи чрез избор на по-съвършена целева функция или чрез използване на дикриминативно обучение на моделите.

[Б3]. A. Lazaridis, I. Mporas, **T. Ganchev**, G. Kokkinakis, N. Fakotakis: *Improving Phone Duration Modelling Using Support Vector Regression Fusion*, Speech Communication, ISSN 0167-6393, vol. 53, no.1, Jan. 2011, pp. 85-97. **(ISI Thomson IF 2012=1.283)**

Постигането на високо качество при синтез на реч зависи от разбираемостта и естествеността и на звучене на синтезирана реч. Прецизното моделиране на продължителността на фонемите е необходимо условие за постигане на естествено звучене на синтезираната реч. Методите за моделиране на продължителността най-общо се разделят на (i) методи основаващи се на правила (rule-based methods) и (ii) статистически методи, които изграждат модел автоматично, след обработка на речта от дадена налична база данни (data-driven methods). Понастоящем статистическите методи за моделиране на продължителността на фонемите са предпочитани, защото те избягват трудоемкото ръчно определяне правила (от експерти по фонетика), което е характерно за други методи. Статистическите методи изграждат модели чрез обработка на големи бази данни, и като цяло не винаги осигуряват необходимата прецизност, поради което усъвършенстването им е особено важно за практиката.

- Статията предлага метод за предсказване на продължителността на фонемите, който се основава на комбиниране на резултатите от множество разнородни предсказатели, които оперират върху общ входен набор от дескриптори. Този метод е основан на наблюдението, че предсказателите изпълнени по разнородни алгоритми имат различно поведение и допускат разнородни грешки. Основавайки се на допускането, че чрез подбиране на подходяща комбинация от изходите на различните предсказатели, грешките биха се компенсирали, което би довело до по-висока точност на предсказанията в сравнение с най-успешният единичен предсказател, бе изследвана различни реализации на метода за комбиниране на резултатите от отделните предсказатели. В тази връзка бе изследвана приложимостта на осем реализации на индивидуални предсказатели, и техните предсказания се използват като входни данни към алгоритъм за комбиниране на резултатите. Изследвани са 12 различни регресивни алгоритъма за реализиране на такова комбиниране: линейна и нелинейна регресия, регресия с невронни мрежи, регресия с опорни вектори, с мета-алгоритми и др. Изследването е проведено за различна степен на гранулация: гласни / съгласни, фонетични категории и индивидуални фонеми.
- Експерименталните резултати с базата данни KED TIMIT (американски английски език) и с базата данни WCL-1 (съвременен гръцки език) показват, че регресия с опорни вектори (SVM регресия) е най-подходящият метод за индивидуален предсказател. Този вид реализация води до намаляване на средната абсолютна грешка с 5.5 %, спрямо втория най-добър метод за база данни KED TIMIT и с 6.8 % за база данни WCL-1.
- Експерименталните резултати показват, че регресия с опорни вектори (SVM регресия) е най-подходящият метод за комбиниране на резултатите от индивидуалните предсказатели. Този вид реализация води до намаляване на средната абсолютна грешка с 1.9 %, спрямо най-добрият индивидуален предсказател, за база данни KED TIMIT, и с 2.6% за база данни WCL-1. По този начин, експерименталните резултати потвърждават работоспособността на предложеният метод и показват, че предложената схема за комбиниране на изходите на индивидуалните предсказатели посредством SVM регресия води до подобряване на прецизността на предсказване на продължителността на фонемите.
- Комбиниране на изходите на набор от разнородни предсказатели посредством SVM регресия е новост при предсказване на продължителността на фонемите.

[Б4]. I. Mporas, **T. Ganchev**, O. Kocsis, N. Fakotakis, O. Jahn, K. Riede: *Integration of Temporal Contextual Information for Robust Acoustic Recognition of Bird Species from Real-Field Data*, International Journal of Intelligent Systems and Applications, ISSN: 2074-904X, 2013.

Запазването на биоразнообразието зависи от наличието на автоматизирани средства за мониторинг, които да подпомагат дейностите по консервация. В частност е особено важно да се наблюдават застрашените животински видове и ключови за екосистемите видове като птиците. Експерти орнитолози инвестират значителни усилия за да проследяват миграциите и поведението на птиците, но традиционните методи са скъпи и ограничени до периодични експедиции, които не могат да установят навременно цялостното развитие на събитията, особено в области с ограничен достъп и в негостоприемни местообитания. Поради тези особености, разработването на технологии за автоматизирано наблюдение на биоразнообразието са необходимост. Към момента има разработки целящи създаването на системи за наблюдение на морски бозайници, птици, прилепи и други видове използващи звуци за комуникация.

- В статията се представя метод за автоматично разпознаване на птици с повишена устойчивост, при който традиционната стратегия за обработка на аудио сигнали е усъвършенствана и допълнена с несложен но ефективен механизъм за интегриране на контекстуална информация. Отчитането на контекста води до подобряване на надеждността на решенията които системата взима. Тук това е реализирано чрез съвместната обработка на резултатите от разпознаването за последователност от няколко аудио сегмента цели отстраняването на спорадични грешни решения, дължащи се на временно влошаване на качеството на сигнала, поради адитивен шум от околната среда.
- В най-елементарния вариант на интегрирането на контекстуална информация се представя като правило, при което ако N предишни и N следващи сегмента са разпознати с добра увереност, че принадлежат на даден вид, то текущия сегмент също се приема за принадлежащ към този вид, независимо от това каква вероятност е изчислена. Големината на прозореца w=2N+1, в която се отчита контекста, зависи от конкретното изпълнение на системата за разпознаване.
- Практическата полезност на предложения метод се оценява на задачата за разпознаване на птици по техните акустични емисии. Експериментите включват изследване на шест широко използвани метода за класификация и база данни от записи на два биологични вида птици, обитаващи планината Имитос край Атина. Записите в базата данни са осигурени от автономни записващи устройства поставени в планината. За да се оцени приложимостта на предложения метод, бяха тествани осем различни метода за класификация, включващи метод на най-близките k-съседа, невронни мрежи, дървета за класификация, класификация базирана на опорни вектори, мрежа на Бейс, и др.
- Предложеният метод постига прецизност при разпознаване от приблизително 93 %, което е добър резултат като се има в предвид, че експериментите са направени с база данни записана в реални, неконтролирани условия. Добавянето на адитивен шум показа значителната устойчивост на метода в условия на ниско съотношение между сигнал и шум, като най-подходящата дължина на прозореца за интегриране на контекстуална информация е w=3. Във всички експерименти интегрирането на контекстуална информация допринесе за подобряване на цялостната прецизност на разпознаването.

Изследванията са финансирани по проекта AmiBio (LIFE08 NAT/GR/000539) "Automatic Acoustic Monitoring and Inventorying of Biodiversity", съ-финансиран от дирекция LIFE+ на Европейската Комисия.

[Б5]. I. Potamitis, **T. Ganchev**, D. Kododimas: *On Automatic Bioacoustic Detection of Pests: The Cases of Rhynchophorus Ferrugineus and Sitophilus Oryzae*, Journal of Economic Entomology, ISSN 0022-0493, vol.102, no. 4, 2009, pp. 1681–1690. **(ISI Thomson IF 2012=1.600)**

Своевременното откриване на вредители е от голямо значение за запазване на реколтата и при съхраняване на продукти. Традиционните методи за борба с вредителите включва периодично вземане на проби и при наличие на заразяване, реколтата или съхранените продукти се третират с химически препарати. Обаче традиционните методи не осигуряват максимална защита, а освен това са скъпи и не винаги ефективни. Биологични видове, които излъчват звуци за комуникация, или в резултат на предвижване или хранене, могат да бъдат откривани автоматично посредством системи за биоакустичен мониторинг.

- В статията се предлага интегриран подход за биоакустично разпознаване на различни видове вредители, като фокусът е на два вида насекоми от семейство Dryophorinae (Coleoptera: Curculionidae),
  - Rhynchophorus ferrugineus Olivier, (RPW), опасен вредител и унищожител на палмови плантации, и
  - Sitophilus oryzae (L.), (гъгрица, rice weevil, RW), всеизвестен вредител в силозите за съхраняване на зърнени култури.

  Този подход се основава на методи за откриване и автоматично разпознаване на акустични емисии, дължащи се на типични поведения на вредителите: придвижване и хранене. Тези акустични емисии се усилват, филтрират, параметризират и класифицират автоматично посредством съвременни статистически методи за моделиране и разпознаване, най-често посредством портативен компютър.
- В статията се изследва усъвършенстван метод за параметризация на аудио сигнала, при която се използва променлива дължина на сегментация на сигнала. Акустичните дескриптори изчислени за всеки сегмент от сигнала включват: кепстрални коефициенти или друг вид дескриптори изчислени посредством дискретното уейвлет пакет преобразуване, и честотата на основната хармонична съставяща на сигнала. Моделирането и разпознаването на акустичните емисии на тези вредители се извършва посредством смес от Гаусови функции.
- Практическата полезност на предложения подход за откриване на двата вида вредители се илюстрира в експерименти с база данни от акустични емисии, записани в реални в неконтролирани условия от заразени палмови дървета, и в складове със заразени зърнени култури. В лабораторни условия бе постигната изключително висока прецизност на разпознаване на тези вредители, която достигна 99.1 % за Rhynchophorus ferrugineus и 100% за Sitophilus oryzae. Кепстралните коефициенти се оказаха по-подходящи дескриптори в сравнение с дескрипторите изчислени посредством уейвлет пакет преобразуването, което се дължи на спецификата на акустичните емисии. Добавянето на честотата на основната хармонична съставяща на сигнала към дескрипторите подобрява прецизността на разпознаване.
- Тези първоначални резултати са твърде оптимистични, тъй като в условията на реална експлоатация на такава система трябва много внимателно да се подбере мястото където да се поставят пиезоелектричните сонди за звукоснемане, така че да бъдат в близост до огнището на заразата. Тъй като мястото на заразата не е лесно предсказуемо, би трябвало при практическата реализация на такива системи да се използват множество сонди, поставени в най-вероятните места за възникване на заразата.

[Б6]. **T. Ganchev**, I. Potamitis: *Automatic Acoustic Identification of Singing Insects*, Bioacoustics: The International Journal of Animal Sound and its Recording, ISSN 0952-4622, vol. 16, 2007. pp. 281–328. **(ISI Thomson IF 2012=0.889)**

В последното десетилетие се предприемат енергични действия за запазване на биоразнообразието на видовете. В някой случаи тези дейности са подпомагани успешно от технологии за наблюдение на физическите и химическите параметри на местообитанията, и автономни камери и записващи устройства. Все още локализацията и разпознаването на насекоми се извършва от висококвалифицирани експерти, които залавят и разпознават насекомите "ръчно", чрез сравнение с налични екземпляри в колекции или техни описания в специализираната литература. В повечето случаи това е трудоемка, бавна и скъпа процедура, особено за насекоми които обитават труднодостъпни места или които се характеризират с предимно нощна активност. Особено важна стъпка за подобряване на ефективността на работа при оценяване на биоразнообразието в дадено местообитание е разработването на технологии за автоматична локализация и разпознаване на биологичните видове посредством методи за аудио и видео наблюдение.

- В статията се предлага метод за параметризация на аудио сигнали с променлива дължина на сегментите, създаден специално за нуждите на автоматизираното разпознаване на насекоми. Предложена е йерархична схема за статистическо моделиране и разпознаване на множество биологични видове по техните акустични емисии. При йерархичната схема класификацията на неизвестен входен сигнал, се извършва след неголям брой сравнения с модели създадени на различни таксономични нива, което позволява да се извлече полезна информация, даже ако входен сигнал е от биологичен вид който не е моделиран от системата за разпознаване. Моделирането използва съставни статистически модели, които се състоят от два разнородни модела (смес от Гаусови функции(GMM) и вероятностна невронна мрежа (PNN)), обучени от общ набор от данни, чиито изходи се комбинират за подобряване на прецизността на разпознаване.
- Сравнена е приложимостта на различни видове дескриптори на сигнала. Предложеният йерархичен метод за разпознаване на насекоми е тестван за 307 биологични вида (щурци, скакалци, цикади) принадлежащи към различни биологични фамилии. Йерархичната схема е сравнена с метод за директно разпознаване, където входният сигнал се оценява последователно за всеки от 307-те модела. За провеждане на експериментите са използвани акустични записи от няколко добре документирани американски, европейски и японски база данни.
- Експерименталните резултати показват че параметризацията на сигнала чрез сегментиране с променлива продължителност е по-успешна отколкото използване на сегменти с предопределена неизменна дължина. Сравнението между трите схеми за класификация (йерархична, йерархична само с две нива, директна), показват че йерархичната схема е най-удачна, понеже намалява грешката от класификация и броя на изчисленията. Това се дължи на заложената стратегия за разделяне на труден проблем на части, които след това се решават една по една, като се разделят на по-малки части. Това позволява да се постигне по-висока цялостна успеваемост на разпознаването, и ниска успеваемост само за няколко много случая, както случая с разпознаването на вида и рода за екземплярите от семейство Tettigoniidae. Обобщавайки резултатите, може да се заключи, че по-голямата сложност при създаване на предложената цялостна схема за параметризиране на сигнала, създаване и обучаване на моделите се отплаща с голямо подобрение в прецизността при разпознаване на голям брой видове.

II. Глави в книги, общо 6 броя

[Б7]. **Ganchev T.** *Locally Recurrent Neural Networks and Their Applications*, Chapter IX in Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods and Techniques, ISBN 978-1-60566-766-9, (Eds: E. Soria, J.D. Martin, R. Magdalena, M. Martinez, and A.J. Serrano.), IGI Global, August 2009, pp.195-222.

Появата и развитието на локално рекурентните архитектури допринесе за значителният напредък в теорията и практическото приложение на невронните мрежи. За кратък период след дефинирането на концепцията за локална рекурентност, бе наблюдавано бурно развитие и поява на множество производни архитектури създадени за нуждите на различни задачи и приложения. Локално рекурентните архитектури станаха важни градивни блокове в многобройни оригинални по замисъл и изпълнение невронни мрежи, които се оказаха изключително успешни при решаване на широк кръг приложни задачи. Общото между тези приложения е изискването за моделиране и разпознаване на пространствени или времеви корелации във времеви редици и структурирани данни. В същото време, многообразието от локално рекурентни архитектури, различните изходни гледни точки при описанието и дефинирането им, както и разнородността на приложенията, в значителна степен затрудняват цялостния поглед и избора на конкретна архитектура подходяща за даден тип практически задачи.

- В тази глава, се предлага унифицирано представяне на всички основни видове локално рекурентни архитектури, от най-елементарните към най-сложните, като се анализират архитектурните особености и различия. В обем от 28 страници читателя се запознава с еволюцията на основните локално рекурентни архитектури и най-характерните им приложения. Унифицираното разглеждане на изчислителните модели за всички структури допринася за по-пълно разбиране на свойствата и възможностите им.
- Предложен е сравнителен анализ на изчислителните модели и ресурсите необходими при реализациите на всяка архитектура.
- Разгледани са типични невронни мрежи базирани на тези архитектури, както и обобщенията им. За типовите невронни мрежи са разгледани примерни решения на практически задачи докладвани в специализираната научна литература. Тези приложения са свързани с класификация или предсказване на времеви редици, откриване и моделиране на пространствени и времеви корелации, идентификация на процеси и управление и др. Обобщените експериментални резултати, докладвани в анализираните публикации, позволяват да се направи извода че локално рекурентните невронни мрежи позволяват да се моделира и открие контекста в който се случват събитията или в който са разположени обектите, което всъщност е причината за преимуществото на локално рекурентните архитектури пред техните нерекурентни съответни архитектури и други конкурентни техники.
- Подчертани са основните характеристики и предимства на локално рекурентните невронни мрежи.
- Очертани са очакванията и насоките за бъдещо развитие на локално рекурентните архитектури и невронни мрежи, включително възможността за ефективната им апаратна реализация посредством новосъздадения четвърти основен пасивен електронен елемент: мемристора.
- За удобство на любознателния читател, в епилога се препоръчват някой Интернет ресурси и форуми, където може да бъде получена допълнителна информация или да бъдат зададени въпроси към специалисти в областта.

[Б8]. **Ganchev T.**, Parsopoulos K.E., Vrahatis M.N., Fakotakis N.: *Partially Connected Locally Recurrent Probabilistic Neural Networks*, Chapter 18 in Recurrent Neural Networks, ISBN 978-3-902613-28-8, September 2008, (Eds: X.-L. Hu and P. Balasubramaniam), InTech, Vienna, Austria, pp.377-400.

Последното десетилетие бележи възраждане на интереса към локално рекурентните невронни мрежи, като се наблюдава задълбочаване на теоретичните изследвания и разширяване на кръга от практическите им приложения. Въпреки значителните усилия за развитие и подобрение на локално рекурентните невронни мрежи, потенциалът за тяхното усъвършенстване далеч не е изчерпан, което се доказва от множеството нови разработки и новосъздадени архитектури, както и от ново-появилите се методи за тяхното обучение. Една от линиите за развитие е създаването на невронни мрежи които имат сходство с биологичните невронни мрежи в човешкия мозък. Развитието на неврологията допринася за по-дълбокото разбиране на архитектурата и функциите на отделните части на човешкия мозък, което стимулира стремежа за създаването на изкуствени невронни мрежи, които наподобяват рекурентните структури на мозъка и отчасти функционирането им.

- В тази глава се дефинира нова архитектура на изкуствена невронна мрежа и метод за обучението и, която обединява концепцията за локално рекурентната вероятностна невронна мрежа (LRPNN) и концепцията за груповия интелект на рояк от индивиди/частици (PSO). Новата архитектура, наречена частично-свързана локално рекурентна вероятностна невронна мрежа (PC-LRPNN), надгражда предишни разработки на авторите и е въведена като обобщение на концепцията за локално рекурентни вероятностни невронни мрежи, с цел приближаването им към биологични структури в кората на човешкия мозък. За разлика от локално рекурентните вероятностни невронни мрежи (LRPNN), при които всички рекурентни неврони са симетрично свързани с всички останали неврони от същия слой, новата PC-LRPNN включва групи (рояк) от свързани неврони в някаква околност, които комуникират помежду си, но които може и да не са свързани с останалите неврони от рекурентния слой.
- Новата PC-LRPNN предлага архитектурно решение, което позволява формирането на рояци от неврони в някаква локална околност и схема за комуникация между тях, която напомня схемите за комуникация предложени в специализираната литература по алгоритми за оптимизация с рояк частици. По принцип топологията на локалната област и размера на рояците могат да бъдат статични или динамични. Пластичността на невронната мрежа е удобна в случаите когато се изисква добро моделиране на недобре проучен процес или непознати явления. В тази глава обект на дискусията е статичната топология, която се дефинира по време на обучението и остава неизменна по време на нормалната работа на PC-LRPNN.
- Използването на връзки и комуникацията между невроните в локална околност, създава условия за оптимизация на връзките в рекурентния слой, което води до много по-бърза работа на PC-LRPNN в сравнение с LRPNN. Обучението и откриването на оптималната за дадена задача топология се извършва с помощта на метод за обучение базиран на алгоритъма за оптимизация с рояк от частици (PSO).
- Ефективността на комбинацията от предложената PC-LRPNN архитектура и метод за обучение с PSO е демонстрирана в две практически задачи: разпознаване на диктори и разпознаване на емоционални състояния, като и в двата случая PC-LRPNN показва превъзходство в прецизността, пред популярната вероятностна невронна мрежа.

[Б9]. **T. Ganchev**, A. Lazaridis, I. Mporas, N. Fakotakis: "*Performance Evaluation for Voice Conversion Systems*", LNAI 5246, Sojka et al. (Eds), ISSN 0302-9743, Springer-Verlag, 2008. pp. 317–324.

Преобразуването на глас е автоматичната модификация на гласа на един диктор така че да звучи като произнесена от друг диктор. Преобразуването на глас намира приложение в системите за синтез на реч, а също може да се комбинира с технологии за засекретяване на комуникации. Процесът за автоматизирано преобразуване на глас се основава на създаване на набор от правила за преобразуване от глас "А" към глас "В". В последствие тези правила се използват за да се преобразува реч от диктор "А", така че да звучи произнесена като от диктор "В". Създадени са множество методи за преобразуване на глас, при което възниква необходимост от тяхното оценяване и сравняване. За целта се използват различни субективни и обективни тестове за оценка на качествата на системите за преобразуване на глас, които оценяват качеството на преобразуваната реч или способността на дадена система да трансформира глас "А" в глас "В". Субективните методи използват група тренирани слушатели, които се произнасят за качествата на системата -- този подход е бавен и скъп, а обективните методи използват мерки за близост, които не отчитат начина по който речта се възприема от слушателя или пък използваните мерки не са интуитивно разбираеми.

- В тази глава се предлага нов метод за обективно оценяване на качеството на системи за преобразуване на глас, наречена "обективен ABX тест", който се основава на съвременни методи и технологии, широко използвани при автоматизираната идентификация на диктори. Този метод се основава на предложената от авторите нова *обективна нормирана мярка* ($D_{NORM} \in [0, 1]$), за оценка на качествата на системите за преобразуване на глас, която може да се използва за оценка на работоспособността на дадена система или за директно сравнение на разнородни по принцип на действие системи. Предложеният метод следва логиката на субективния ABX тест, при който тренирани слушатели оценяват субективно дали дадена фраза или изречение "X", получена от система за преобразуване на глас, звучи по-близо до първоначалния глас "А" или до желания глас "В".
- Предложеният метод за обективно оценяване на качеството на системи за преобразуване на глас използва известно количество записана реч от глас "А" и глас "В" за да създаде статистически модели на гласовете на двата диктора. След което, речта получена от системата за преобразуване на глас се сравнява с всеки от моделите и се нормализира спрямо универсален модел на речта. В резултат на нормализацията се получава обективна нормирана мярка за близост, която е интуитивна, лесно разбираема и удобна както за числово така и за графично представяне. В конкретната разработка, обективната нормирана мярка се изчислява посредством GMM-UBM оценител, който е добре познат като основен градивен елемент на съвременните системи за идентификация на диктори. GMM-UBM оценителя използва три модела: универсален модел на речта, и модели "А" и "В" получени след максимално-апостериорна адаптация с използване на записана реч от диктори "А" и "В".
- Работата и възможностите на предложеният метод са представени нагледно чрез сравнение на две класически системи за преобразуване на глас. Предложената нова обективна нормирана мярка D_{NORM} е сравнена с 4 добре познати обективни критерия за близост.

Изследванията са финансирани по международният проект PlayMancer (FP7 215839) "A European Serious Gaming 3D Environment", съфинансиран по седма рамкова програма (FP7) на Европейската Комисия.

[Б10]. A. Lazaridis, **T. Ganchev**, I. Mporas, T. Kostoulas, N. Fakotakis: "*Feature Selection for Improved Phone Duration Modeling of Greek Emotional Speech*", S. Konstantopoulos et al. (Eds.): Advances in Artificial Intelligence: Theories, Models, and Applications, LNAI 6040, Springer-Verlag Berlin Heidelberg 2010, pp. 357–362.

Синтезът на емоционална реч, необходим за нуждите на съвременните интерфейси човек-машина, се основава на информация извлечена от естествена човешка реч. От особено значение за постигане на висококачествен синтез на реч е изясняване влиянието на различни прозодични характеристики, важни при възприятието на емоционална реч от човека. Прозодията има важна роля в комуникацията между хората, тъй като носи важна информация, която може да не е представена (или да не е еднозначно формулирана) в граматичната конструкция на изказа. Прозодията позволява да се покаже подчертаване на определен смисъл, намерение или емоционално състояние, което има значение за цялостното възприемане и разбиране смисъла на съобщението.

Един от основните фактори, които допринася за генерирането на високо-качествена синтетична емоционална реч, е прецизно моделиране на дължината на фонемите при различните емоционални изкази.

- С помощта на широко използваните методи за подбор на корелирани дескриптори, известни като Relief и корелативен метод за избор на описатели (Correlation based Feature Selection, CFS) е изследвана възможността за подобряване на прецизността при моделиране на продължителността на фонемите. Хипотезата е, че чрез подбор на подмножество от дескриптори на речта, които са силно корелирани с продължителността на фонемите, и в същото време изключване на дескриптори, които не допринасят значително за подобряване на прецизността на моделите, би подобрило прецизността на определяне на продължителността на фонемите.
- В изследването са използвани 93 дескриптора на речта, 4 категории емоционална реч (гняв, страх, печал, радост) и емоционално неутрална реч и 10 различни метода за моделиране на продължителността на фонемите, използващи: дървета на решенията, линейна регресия, алгоритми с „мързеливо" обучение и мета-алгоритми.
- Експериментите са извършени с използване на база данни от емоционална реч на съвременен гръцки език, която съдържа думи, фрази и изречения с неутрални лингвистично съдържание, но с емоционален изказ, пресъздаден от професионална актриса.
- Експерименталните резултати показват, че използването на селективен подбор на дескрипторите води до подобряване на прецизността на предсказване на продължителността на фонемите при емоционално неутрална реч и при всички категории емоционална реч, с изключение на категорията „радост". Независимо кой от 10-те метода за моделиране е използван, селективният подбор на дескриптори намалява средно квадратичната грешка на предсказване на продължителността на фонемите с до 2 %, в сравнение с тази получена от най-прецизният модел използващ всичките 93 дескриптора. От друга страна използването на селективен подбор на дескрипторите доведе до подобряване на корелационният коефициент на използваните подмножества от дескриптори с до 3%.
- Експерименталните резултати потвърждават хипотезата, че изключване на дескриптори които са слабо корелирани с продължителността на фонемите, и по този начин допринасящи слабо за подобряване на модела, оказва благоприятен ефект върху прецизността на предсказанията, независимо от метода за моделиране и предсказване на продължителността на фонемите при емоционална реч. В допълнение се подобрява компактността на модела и се понижават изискванията за изчислителна мощ.

[Б11]. **T. Ganchev**, I. Mporas, O. Jahn, K. Riede, K.-L. Schuchmann, N. Fakotakis: "*Acoustic Bird Activity Detection on Real-Field Data*", I. Maglogiannis, V. Plagianakos, and I. Vlahavas (Eds.): SETN 2012, LNAI 7297, Springer-Verlag Berlin Heidelberg, 2012. pp. 190–197.

Птиците са важен индикатор за състоянието на местообитанията и естествените пейзажи. Откриване присъствието на определени видове птици, оценяване тенденциите за изменение на популациите им и успешността при възпроизводството им, е от изключителна важност при оценяване състоянието на екосистемите. Понастоящем наблюдението на биоразнообразието и преброяването на наличностите от птици се извършва от експерти орнитолози, които периодично посещават местообитанията и провеждат аудио-визуални наблюдения, или използват методи като улавяне-пускане-улавяне за оценка на популациите. Използването на експерти е скъпо и трудоемко, поради което е трудно да се организира непрекъснато наблюдение за дълги периоди от време, особено в трудни климатични условия. По тези причини, даже частична автоматизация на процеса би подпомогнала значително усилията за оценка на биоразнообразието на видовете.

- В тази глава се докладва разработването на автоматичен детектор за откриване на акустична активност на птици (ABAD), който е важна част от системите за автоматизирано разпознаване на птици по техните акустични емисии. Детекторът използва база данни от записи на птици събрана в планината Имитус край Атина. Част от базата данни е щателно анотирана от експерт орнитолог за нуждите на научно-изследователската и развойна дейност по проекта AmiBio. Тук, анотираните записи са използвани за създаване на статистически модел за откриване на птичи вокализации сред всички останали звуци в природата.
- Експерименталното изследване бе насочено към оценяване на прецизността на разпознаване на птичи звуци в реално време, поради което системата бе настроена да взема решения за краткотрайни 20 ms отрязъци от аудио на всеки 10 ms. Това води до песимистична оценка на работоспособността на детектора, която в случая е 86% коректни решения.
- Изследвани са различни методи за машинно обучение за нуждите на статистическото моделиране на птичите вокализации, като целта е да се подбере метод който да е подходящ за реализация на детектора в преносимо автономно устройство, разполагащо с ограничени ресурси на батерията и ограничена памет и изчислителна мощ. Изследвани са пет метода за моделиране, измежду които метод на най-близките k-съседа, трислойна невронна мрежа с архитектура 18-10-1 и сигмоидални неврони, дърво на решенията C4.5, класификатор базиран на опорни вектори и ядро с радиални базисни функции, и Бейсова мрежа. Всички методи бяха изследвани по общ експериментален протокол и еднакви условия.
- Експерименталните резултати показаха най-висока прецизност (измежду петте тествани метода) при използване на трислойната невронна мрежа с архитектура 18-10-1. Отчасти този резултат се дължи на малкият набор от тренировъчни данни използван за изграждането на статистическите модели.
- Способността на трислойната невронна мрежа с 10 неврони в скрития слой да осигурява добра прецизност я прави подходяща, тъй като осигурява компактно и изчислително евтино решение, което е удобно за реализация в преносими устройства с ограничени енергийни и изчислителни ресурси.

Изследванията са финансирани по проекта AmiBio (LIFE08 NAT/GR/000539) "Automatic Acoustic Monitoring and Inventorying of Biodiversity", съ-финансиран от дирекция LIFE+ на Европейската Комисия.

[Б12]. I. Mporas, **T. Ganchev**, O. Kocsis, N. Fakotakis, O. Jahn, K. Riede, K.-L. Schuchmann: "*Automated Acoustic Classification of Bird Species from Real -Field Recordings*", Proc. of the IEEE International Conference on Tools with Artificial Intelligence (ICTAI-2012), November 7-9, 2012, Athens, Greece. Vol.1, pp.778-781.

Наблюдението на птичите популации е изключително важно за запазването на биоразнообразието в екосистемите. Традиционните методи за наблюдение от експерти орнитолози са скъпи и трудоемки, поради което не могат да отговорят на нуждите за непрекъснато наблюдение на различни местообитания и екосистеми. През последното десетилетие се появиха системи автоматично разпознаване на птичите видове по техните вокализации, които могат да разпознават ограничен брой птичи видове. Тези системи най-често разпознават само един или няколко вида птици и работят устойчиво в лабораторни условия, но не се справят добре в реални условия, тъй като присъствието на шум и звуци от конкурентни източници влошава значително прецизността на разпознаване.

- Докладът представят резултати от усилията на колектива да създаде автоматична система за разпознаване на птичи видове по техните вокализации, необходима за нуждите на проекта AmiBio[1]. Системата е базирана на класическия подход за обработка на аудио сигнали, която е намира приложение при разпознаване на аудио събития. Първоначално, аудио сигналът се подава към детектор на птичи вокализации, които пропуска за по-нататъшна обработка само сегментите разпознати като "птичи сигнал/песен" и елиминира всички останали звуци. Следващото стъпало класифицира отрязъците разпознати като "птичи сигнал/песен" към един от познатите модели или определя отрязъка като неразпознат. В доклада този подход е тестван за седем вида птици, които най-често се срещат в планината Имитус.
- В доклада се изследва работоспособността на модула за класификация на седем вида птици. Изследвани са шест различни метода за реализация на класификатора (AdaBoost(J48), Bagging(J48), Баесова мрежа, класификатор с опорни вектори, най-близки к-съседа и дърво на решенията J48), като експеримента включва използване на записи направени в реални условия в планината Имитус. Най-висока прецизност от около 93 % е постигната с метода AdaBoost(J48), следван от Bagging(J48) с около 92 %.
- В допълнение към експериментите с реални записи, така както за записани е планината Имитус, бяха извършени допълнителни изследвания на шумо-устойчивостта на системата. За целта, бе добавян адитивен шум за реализиране на ниски нива на отношението сигнал-шум: -6 dB, 0 dB, 6dB, 12 dB[2], при което бе наблюдавано умерено спадане на прецизността на класификатора с около 17 % (прецизност в абсолютни единици).

Изследванията са финансирани по проекта AmiBio (LIFE08 NAT/GR/000539) "Automatic Acoustic Monitoring and Inventorying of Biodiversity", съ-финансиран от дирекция LIFE+ на Европейската Комисия.

[1] Дългосрочната цел на проекта AmiBio е да установи автоматизирана система за наблюдение на биоразнообразието в планината Имитус край Атина. Пилотния етап е завършен през Юни 2013.

[2] Трябва да се отбележи, че отношението сигнал-шум е измервано за отрязъци от аудио записа, така че конкретно за интервалите от сигнала, в които има открита птича вокализация, е възможно отношението сигнал шум да е по-високо от средното.

III. Рецензирани доклади от конференции в сборник доклади (8 броя)

[Б13]. R. Saeidi, **T. Ganchev**, H.S. Mohammadi: "*Text independent speaker verification using enhanced sorted Gaussian mixture model*", Proc. of the 2007 IEEE International Conference on Signal Processing and Communication (ICSPC-07), Dubai, United Arab Emirates (UAE), November 24-27, 2007. pp.1191-1194.

Съвременните системи за разпознаване на диктори използват смеси от Гаусови функции (GMM) за статистическо моделиране на разпределението на даден вид дескриптори на речта. Най-устойчиво моделиране се получава по метода UBM-GMM, при който индивидуалните GMM модели на дикторите се получават чрез адаптиране на един общ модел на речта (UBM), който е създаден от речта от голям брой диктори. За извършване на адаптирането е необходимо известно количество реч от целевия диктор, което се използва за донастройка на параметрите на индивидуалния GMM модел, най-често по метода на максималната апостериорна адаптация. При тази процедура се получава GMM модел на диктора, който е близко свързан с универсалния модел (UBM), тъй-като всеки компонент на GMM е получен чрез донастройка на UBM. За да се намали броя на изчисленията при верификация на диктори са предложени различни икономични схеми за определяне степента на близост между двойката UBM-GMM и даден входен сигнал. Една от тези схеми реализира предложения от авторите метод за сортирана смес от Гаусови функции (SGMM). Сортираната смес от Гаусови функции многократно намалява обема на изчисленията при изчисляване на резултата за даден входен сигнал.

- В доклада се предлага усъвършенствана сортираща функция, която подобрява работата на предложения по-рано метод със сортирани смеси от Гаусови функции (SGMM). Усъвършенстваната сортираща функция позволява да се получи още по-голямо съкращаване на обема на изчисленията, в сравнение с оригиналната сортираща функция използвана в първоначалната формулировка на метода.
- Експерименталното сравнение на двете сортиращи функции е проведено с помощта на две разнородни бази данни: първата съдържаща висококачествена реч на персийски език (фарси) записана от телевизионни програми а втората 2002 NIST SRE One-speaker detection data съдържаща телефонна реч на английски език.
- Експерименталните резултати от базата данни с висококачествена реч показват, че усъвършенствана сортираща функция дава по-добри резултати от оригиналната, най-вече в случаите където се залага на многократно намаляване на обема на изчисленията и когато данните с които са създадени моделите съвпадат с оперативните условия в които системата за верификация работи. Когато има несъответствие между акустичната среда в която е събрана речта за обучение на моделите и средата в която работи системата за верификация на диктори, както е при базата данни 2002 NIST SRE, оригиналната сортираща функция дава по-добри резултати. Това наблюдение подсказва че оригиналната сортираща функция е по-малко чувствителна към несъответствия между акустичните условия в които системата е обучена и в които работи. Друга съществена разлика между двете бази данни е дължината на тестовите записи. Изглежда че оригиналната сортираща функция по-добре използва по-дългите записи с реч на диктора, които в 2002 NIST SRE достигат до повече от минута за някой тестове.
- Използването на сортирана смес от Гаусови функции позволява намаляване на обема на изчисленията 14 пъти, спрямо изчисленията необходими за оригиналния метод с UBM-GMM, при незначително увеличаване на балансираната грешка при верификация с 0.43 % (в абсолютни единици).

[Б14]. T. Kostoulas, **T. Ganchev**, I. Mporas, N. Fakotakis: *"A Real-World Emotional Speech Corpus for Modern Greek"*, Proc of the LREC-2008, Morocco, May 28-30, 2008.

Навлизането на речеви диалогови системи в ежедневието ни повишава изискванията към потребителския интерфейс. Особено важна стъпка към подобряване на възприятието на потребителите е снабдяване на диалоговите системи с възможността да усещат емоциите на клиента по гласа. Съвременните технологии за разпознаване на емоционални състояния по гласа използват статистически методи, като смеси от Гаусови функции, динамични Бейсови мрежи, и др., които моделират основните (първични) емоции и създават модели, които са представителни за големи групи потребители. За да бъдат създадени тези статистически модели са необходими представителни и балансирани бази данни, които да бъдат надлежно анотирани и сертифицирани от експерти. Тези бази данни трябва да са реалистични спрямо приложението за което ще се изграждат моделите и да съдържат голям брой примери от всяка категория емоционални състояния, които ще се моделират.

- В доклада се представя първия етап от създаването на нова база данни с емоционална реч, проектирана и реализирана от авторите за нуждите на изследователската и развойни дейности по разпознаване на емоционални състояния. Целта на първия етап бе да се събере емоционална реч от най-малко 40 потребителя, които нямат предишен опит с диалогови системи от типа "умен дом". След реализирането първия етап, базата данни съдържа записи от 43 потребителя говорещи гръцки като майчин език, мъже и жени, които са записвани в 25 минутни сесии в условията на "умен дом". Всички сесии бяха записани в студио оборудвано с мебели, телевизор, аудио уредба, компютър и др. за да наподобява домашна обстановка в "умен дом". За по-голяма реалистичност на изказа, потребителите не бяха инструктирани какви команди и словосъчетания да използват, но бяха инструктирани по какъв начин да активират системата и кои са домашните електроуреди, които могат да бъдат управлявани с глас. Работата на домашните уреди беше симулирана с помощта на анимация, прожектирана на голям екран на стената. Всяка сесия е записвана с видеокамера, а гласът на потребителите е записван с близък висококачествен микрофон и с отдалечен масив от 8 микрофона. Видео записа бе използван в помощ на анотирането на емоционалното състояние. Извършена е анотация на всеки запис в базата данни на лингвистично и емоционално ниво. Предложено е балансирано разделение на базата данни на набор за обучение на модели и набор за тестване на модели.
- Проведени са експерименти със съвременна система за разпознаване на емоционални състояния по гласа на потребителя, използваща статистически модели реализирани със смеси от Гаусови функции. Експериментите бяха насочени към оценяване на трудностите при разпознаване на емоции в реалистични приложения, а също и за оценяване на полезността на базата данни. В експериментите са застъпени сценарии където всеки потребител е моделиран поотделно и където се създават модели общи за всички потребители. Получените експериментални резултати разкриват значителната трудности при разпознаването на спонтанни емоции, в сравнение с базите данни с поръчкови емоции изиграни от професионални актьори. Базите данни с поръчкови емоции, изиграни от актьори, обаче са неприложими при създаването на реални системи за разпознаване на емоции в практически приложения.

Проектирането и осъществяването на базата данни е частично подпомогнато от дейностите и изследователските и развойни дейности по международният проект PlayMancer (FP7 215839) "A European Serious Gaming 3D Environment", съфинансиран по седма рамкова програма (FP7) на Европейската Комисия.

[Б15]. T. Kostoulas, O. Kocsis, **T. Ganchev**, F. Fernández-Aranda, J.J. Santamaría, S. Jiménez-Murcia, M. Ben Moussa, N. Magnenat-Thalmann, N. Fakotakis, "*The PlayMancer Database: A Multimodal Affect Database in Support of Research and Development Activities in Serious Game Environment*", Proc. of the Seventh conference on International Language Resources and Evaluation (LREC'10), May 19-21, Valletta, Malta, Eds. N. Calzolari et al., European Language Resources Association (ELRA), ISBN: 2-9517408-6-7. pp. 3011-3015.

В последните години в допълнение към видео игрите създадени с единствена цел развлечение, се появи нова категория видео игри, наречени сериозни игри. Целта на сериозните игри най-често е подпомагане на обучението в училище, тренировки за военнослужещи, за подобряване на самоуважението и психическата устойчивост на деца, подрастващи и възрастни, за подобряване на информираността за някой болести, за подобряване на дисциплината и придържането към определена терапия, за подобряване на уменията за решаване на проблеми и др. Изследва се използването на сериозни игри при терапия на пациенти с душевни заболявания и разстройства, включително депресии, злоупотреба с алкохол, след травматичен стрес и др. В научната литература е документирано, че някой видео игри могат да подобрят настроението и намалят стреса. Създаването на такива видео игри изисква разработването на интуитивни потребителски интерфейси, които да улавят емоционалното състояние на потребителя. Технологиите за разпознаване на емоционалното състояние на потребителя обикновено използват статистически методи за моделиране, които от своя страна се нуждаят от представителни за приложението бази данни.

- В доклада се представя нова мулти-модална база данни и метода по която е създадена. Предназначението на базата данни е подпомагане на изследователските и развойни дейности по разпознаване на емоционални състояния за нуждите на проекта PlayMancer. Тези дейности са насочени към създаването на сериозни игри, подпомагащи лечението на пациенти с поведенчески разстройства или пристрастяване, и по конкретно разстройства на храненето и пристрастяване към хазарта.
- Базата данни се състои от 18 сесии, всяка с продължителност около 1 час, съдържащи синхронизирани записи на някой жизнени показатели и реч от здрави доброволци, мъже и жени на възраст между 18 и 43 години, предимно от Испания, които са записани по време на занимания с видео игри. Всеки запис включва: реч, видео, биосигнали като честота на пулса, насищане на кръвта с кислород и др., като по време на сесията участника бе оставен сам в стаята, за да се избегне влиянието от ефекта на наблюдавания.
- За нуждите на проекта от особен интерес са както основните емоции: гняв, отегчение, радост, неутрално, тъга, изненада, така и емоционалните състояния с висока интензивност на гняв и униние, така че анотациите бяха насочени към тези емоции. Всеки запис е анотиран от самия участник в експеримента (self-annotation method), тъй като той/тя наблюдавайки записа може най-добре да оцени емоционалното си състояние.
- Въпреки че базата данни е създадена за нуждите на проекта PlayMancer, тя е достатъчно универсална по организация и съдържание, така че би могла успешно да се използва при създаването на мулти-модални интерфейси човек-машина.

Проектирането и осъществяването на базата данни е финансирано по международния проект PlayMancer (FP7 215839) "A European Serious Gaming 3D Environment", съфинансиран по седма рамкова програма (FP7) на Европейската Комисия.

[Б16]. **T. Ganchev**, I. Potamitis, N. Fakotakis: "*Acoustic Monitoring of Singing Insects*", Proc. of the ICASSP-2007. Vol.4, pp.721-724.

Биоакустичната идентификация на насекомите се основава на акустичните емисии, които те излъчват за целите на комуникация или на акустичните емисии дължащи се на хранене, летене или предвижване. Практическата значимост и потенциалните приложения на автоматизираната идентификация на насекоми се дължи на факта, че насекомите имат важно значение в селското и горското стопанства, като някой от тях се явяват вредители по отглежданите култури. Наличието или отсъствието на някой видове насекоми е добър индикатор за нарушения в екосистемата или понижаване на биоразнообразието. Ръчното откриване и разпознаване на насекомите е сложна и скъпа процедура, при която са необходими експертни умения. Освен това е по-лесно да чуеш звуците на насекомо, отколкото да го откриеш и хванеш, особено за тези които обитават сложни местообитания или са активни главно през нощта. Обучението на експерти е сложно и продължително и изисква поддържането на скъпи колекции от насекоми и наличието на специализирана литература. Автоматизираното разпознаване и наблюдение на насекомите, анализ на жизнеността на популациите им, оценка на местообитанията им, ранното откриване на вредители, би създало условия за съхраняване на биоразнообразието в екосистемите.

- В доклада се предлага цялостна концепция на система за автоматичното разпознаване на насекоми по техните емисии, чрез адаптиране на технологии широко използвани в приложения за обработка на реч.
- Предлага се метод за сегментация на аудио сигнали, който води до сегментиране на сигнала на сегменти с променлива дължина. Всеки сегмент, в който е отчетена някаква акустична активност, се полага на последваща параметризация, която е пригодена към особеностите на акустичните емисии на насекомите, за извличане на дескриптори на сигнала. Така получените дескриптори се използват в схема за класификация, която е реализирана посредством съвременни статистически методи за моделиране. Изследвана е приложимостта на три различни метода за класификация: вероятностна навронна мрежа (PNN), смес от Гаусови функции (GMM), и скрити модели на Марков (HMM), които са поставени да работят в условия на недостиг от данни за обучение.
- Експерименталните изследвания включват разпознаването на най-шумните насекоми: щурци, цикади (жътвари), големи американски скакалци. Целта е да се разпознаят фамилиите, под-фамилиите и биологичния вид към който дадено насекомо принадлежи. Експериментите бяха проведени с помощта на голяма базата данни известна като North America collection (SINA), която е анотирана от експерти по таксономия на насекомите.
- Прецизността на предложената система за автоматична идентификация на насекоми надхвърля 98 % за ниво фамилия и достига 86 % на ниво биологичен вид, при разпознаване на 313 биологични вида. Сравнително високата прецизност се дължи на факта че по-голямата част от записаните акустични емисии са с високо качество и/или са почистени внимателно от смущения и конкурентни звуци. Получените резултати са оптимистични и заради факта че половината от всеки запис е използвана за обучение а другата половина за оценка на прецизността. Така в експеримента се получава съвпадане на акустичните условия в които са направени записите за обучение и тестване, което улеснява разпознаването.

[Б17]. I. Potamitis, **T. Ganchev**, N. Fakotakis: "*Automatic Bioacoustic Detection of Rhynchophorus Ferrugineus*", Proc. of the EUSIPCO 2008, August 25-29, Luisiane, Switzerland.

В последното десителетие един опасен вредител, известен като червена палмова гъгрица *Rhynchophorus ferrugineus* (red palm weevel), застрашава палмовите плантации в Средиземноморието и Близкия Изток. Засегнати са палмови плантации в над 30 държави измежду които Испания, Гърция, Египет, Кипър, Израел и др., където вредителя унищожава плантации от кокосови, финикови и сагови палми, плантации за добив на палмово масло, декоративни палми и др. Появата на този вредител обикновено завършва с унищожаване на плантациите, поради което ранното му откриване е от първостепенна важност. Трудността при откриването на заразата идва от това, че ларвите се развиват вътре в дънера, хранейки се го унищожават, и когато достигнат зрялост излитат за да заразят други палмови дървета. По такъв начин вредителя може да се открие визуално, едва след като палмовото дърво е унищожено и възрастните насекоми са излетели за да снесат яйцата си в други дървета. Химическите методи за борба с вредителя са затруднени от това че ларвите живеят вътре в дънера, които ги защитава, а механичните капани за възрастни насекоми не могат да спасят застрашените дървета. Ранната диагностика с акустични методи позволява да се открие заразата в палмови дървета, преди да ларвите да са станали възрастни, което дава шанс да се прекъсне цикъла на заразяване и да се спаси незаразената част от плантацията. Обаче акустичния метод изисква високо-квалифицирани биолози, специалисти по вредители, да прегледат и прослушат всяко едно дърво, което ограничава възможностите за бързо противодействие на вредителя.

- В доклада се предлага метод за автоматичен биоакустичен детектор за ларвите на вредителя *Rhynchophorus ferrugineus*, които позволява ранното откриване на заразата в палмовите плантации. Всъщност, детектора открива пукащите звуци от скъсване на фибрите на дървото, които ларвите прекъсват при хранене. Акустичния сигнал е сравнително слаб, но е възможно да бъде уловен по контактен метод.
- За откриване на акустична активност дължаща се на храненето на ларвите на вредителя, се използват пиезоелектрични сензори, които се поставят в отвори в дънера на дървото. Бе използвана промишлена система за звукоснемане, данните от която се обработват по предложения в доклада метод. Накратко, получения сигнал се подлага на цифрова обработка (с помощта на преносим компютър), където се усилва, филтрира, сегментира, параметризира. Получените дескриптори се използват за моделиране и впоследствие за автоматична класификация при която входния сигнал се сравнява с предварително създадения модел на звуковия профил на вредителя. Предложения метод позволява множество палмови дървета да бъдат следени непрекъснато и при наличие на зараза да бъде изпратено предупреждение до собственика на плантацията.
- В експерименталните изследвания е използвана база данни от записи на акустична активност на ларвите на *Rhynchophorus ferrugineus*, направени от реално заразени палмови дървета в палмовата гора Вай на о. Крит. Записите са анотирани от експерти биолози.
- Експерименталните резултати получени при обработка на базата данни от записи на *Rhynchophorus ferrugineus* събрана в реални условия, показаха че предложения автоматичен метод е достатъчно прецизен за надеждното разпознаване на заразата. Прецизността на откриване на вредителя достига до около 99.5 % при експериментите в лабораторни условия.

[Б18]. I. Potamitis, **T. Ganchev**, N. Fakotakis: "*Automatic Acoustic Identification of Insects Inspired by the Speaker Recognition Paradigm*", Proc. of the InterSpeech-2006 -- ICSLP, Pittsburgh, PA, USA, Sept 17-21, 2006, pp.2126-2129.

Напредъка на компютърните технологии и развитието на методите за цифрова обработка на сигнали и машинно обучение създаде предпоставки за разработване на системи за автоматична идентификация на биологичните видове посредством видео и акустично наблюдение. Все още обаче, адаптирането на тези методи към спецификата на идентификацията на биологичните видовете е в начален етап, поради което не са достигнати надеждни решения.

- В доклада се предлага метод за биоакустична идентификация на видовете, които се базира на технологии които първоначално са разработени за нуждите на разпознаване на реч, и в последствие са успешно адаптирани за разпознаване на диктори. Предложеният метод за класификация на акустични записи на насекоми до категории, като фамилии, под-фамилии и биологичен вид, използва моделиране със смес от Гаусови функции, т.е. статистически метод за машинно обучение широко използван при идентификацията на диктори. Обаче поради значителните разлики в механизма на звуко-образуване при насекомите и човека, предложеният в доклада метод значително се различава от методите за параметризация на сигнала използвани при обработка на речеви сигнали.
- Предложени са два метода за определяне на дескриптори на акустичните емисии на насекоми и е изследвана тяхната приложимост за биоакустична идентификация на насекоми. Първият метод използва сегментация с неизменна дължина на прозоречната функция, и припокриващи се сегменти от сигнала. За всеки сегмент от сигнала се определят кепстрални коефициенти (LFCC), при които мел-банката филтри е заменена с линейна банка филтри. Вторият метод използва сегментация с променлива дължина на сегмента, при който се откриват и отделят за последваща обработка само участъци в които е открита акустична активност, а участъците съдържащи само фонов шум не се подлагат на параметризация. За участъците, в които е регистрирана акустична активност, се изчисляват: кепстрални коефициенти определени посредством линейна банка филтри (LFCC), честотата на доминиращият хармоник в сигнала и дължината на сегмента.
- Двата метода за параметризация на сигнала са сравнени експериментално. За целта са използвани записи от 105 вида насекоми (щурци и скакалци от Северна Америка), от базата данни Singing Insects of North America -- SINA. Метода за изчисляване на дескрипторите на сигнала посредством променлива дължина на сегмента позволява постигането на по-висока прецизност и по-добро бързодействие в сравнение с метода с неизменна дължина на сегментация. Освен това е показано експериментално, че доминантната честота на сигнала не е достатъчна за надеждна идентификация на биологичните видове, под-фамилии и фамилии. Съвместното използване на доминантната честота на сигнала и дължината на сегмента значително повишава прецизността на класификация, а най-висока прецизност се получава при съвместното използване на трите вида дескриптори на сигнала: LFCC, честотата на доминиращия хармоник в сигнала и дължината на сегмента.
- В експериментите проведени в лабораторни условия е достигната прецизност от над 99 % при разпознаване към коя фамилия (една от три) принадлежи даден запис. На ниво под-фамилия (една от шест) прецизността достига до 98 %, а на ниво биологичен вид (105 вида насекоми) прецизността достига 94 %.

[Б19]. **T. Ganchev**, A. Lazaridis, I. Mporas, N. Fakotakis: "*Average Performance Loss Measure for Assessment of Voice Conversion Systems*", Proc. of the 13-th International Conference on Speech and Computer SPECOM'2009, June 21-25, St. Petersburg, Russia. pp. 275-280.

Съществуват множество обективни и субективни методи за оценка качеството на системите за преобразуване на глас. Всеки от тези методи се основава на предварително избрана мярка за качеството на сигнала за преобразувания глас или за близост между преобразувания сигнал и целевия глас. Субективните методи се основават на слушателски тестове, където тренирани експерти прослушват множество записи за да сравнят качествата на две системи. Това е трудоемка процедура, която изисква висококвалифицирани специалисти, поради което е скъпа. От друга страна обективните тестове не отчитат за качеството на възприятие на преобразувания сигнал, не са интуитивни или са неустойчиви при големи отклонения в качеството на сигнала за трансформирания глас.

- В доклада се анализират недостатъците на най-често използваните методи за оценка на системи за преобразуване на глас и се предлага нов обективен метод, наречен SVT, за оценка на системите за преобразуване на глас. Предложеният метод използва нова и интуитивна мярка за оценка на близостта на преобразувания глас до първоначалния и до целевия глас. Този метод отчита и преодолява много от недостатъците на някой предшестващи го методи.
- Предложеният нов обективен метод следва логиката на субективния ABX метод, но реализира оценяването на системите за преобразуване на глас по различен начин. Предложения метод използва технологията с GMM-UBM оценител, които сравнява входния сигнал с общ модел на речта (UBM) и модел на преобразувания глас (GMM) който се получава от UBM след максимална апостериорна адаптация. За разлика от метода предложен в публикация [Б9], където се съставят модели за първоначалния "A" и целевия глас "B", тук се съставя модел само за преобразувания глас "X", при което всички експерименти се свеждат до верификация на дикторите, "A", "B" и "X". Това улеснява настройката и калибрацията на метода за оценка, и позволява да се дефинира по-гъвкава и по-интуитивна мярка за близост наречена "average performance loss" (ALP), при която APL=0 означава идеално съвпадение между преобразувания и целевия глас, т.е. идеална система за преобразуване. Резултат APL<<1 означава добра система, резултат APL=1 посредствена система и APL>1 лоша система за преобразуване на глас. Предложеният метод е подходящ както за оценяване на качествата на една система, така за сравняване на две системи, които работят на сходен или коренно различен принцип.
- Работа на предложения метод е изследвана и илюстрирана чрез експерименталното сравнение на две разнородни системи за преобразуване на глас. Представени са графики на междинни резултати които улесняват изясняването на логиката на теста.
- Представени са сравнителни експериментални резултати, при които две разнородни системи за преобразуване на глас са оценени посредством предложения метод и 5 други предшестващи метода за оценка. Направен е анализ на предимствата и недостатъците на отделните методи за оценка на системите за преобразуване на глас, като предложената нова мярка и нов метод преодоляват недостатъците на предшестващите методи и показват устойчивост дори при лошо качество на преобразувания сигнал.

Изследванията са финансирани по международният проект PlayMancer (FP7 215839) "A European Serious Gaming 3D Environment", съфинансиран по седма рамкова програма (FP7) на Европейската Комисия.

[Б20]. S. Ntalampiras, **T. Ganchev**, I. Potamitis, N. Fakotakis: "*Objective Comparison of Speech Enhancement Algorithms under real world conditions*", Proc. of the 1st International Conference on PErvasive Technologies Related to Assistive Environments, PETRA-2008, July 15-19, Athens, Greece.

Методите за шумоподтискане, използвани в контекста на обработка на речеви сигнали, целят отстраняване на конкурентни смущения така че да се подобри възможността за възприемане на съобщението от получателя (машина или човек). Когато крайната цел е подобряване на прецизността при разпознаване на реч или на диктори, не се изисква възстановяване на сигнала във времевата област, което дава известна свобода при използването на нелинейни трансформации. Когато целта е подобряване качеството на речта, е необходимо да се избягва внасянето на нелинейни изкривявания в сигнала по време на процеса на шумоподтискане. Разнородният характер на смущенията и голямото разнообразие на условията при експлоатация на мобилните устройства затруднява намирането на метод или решение, което да осигурява надеждно шумоподтискане при всички ситуации. Поради това, обикновено методите за шумоподтискане се разработват за ефективно справяне с определен тип смущения, които преобладава или който създава най-сериозни проблеми.

- В доклада се изследва работата на осем различни метода за шумоподтискане, които се причисляват към следните четири основни категории: (i) методи използващи спектрално изваждане, (ii) методи използващи статистическо моделиране, (iii) методи работещи в подпространството на сигнала, и (iv) методи използващи специални алгоритми за филтрация.
- Целта на изследването е да се оцени пригодността на тези методи за потискане на бързо изменящи се смущения, характерни за движещ се мотоциклет. В изследването се използва базата данни MoveOn съдържаща реч (съответстваща на командния протокол в полицейските сили) и бързо-изменящ се шум записани при патрулиране на полицейски мотоциклети в градски и извън градски условия.
- Подбрани са осем разнородни метода за шумоподтискане, базирани на спектрално изваждане, подпространства на сигнала, модели на шума и сигнала, и Калман филтри, които се сравняват с помощта на субективни и обективни мерки за определяне качеството на сигнала след потискането на смущенията. Използвана е стандартизираната обективна мярка за качеството на сигнала PESQ - ITU-T Rec. P.862, която оценява качеството на сигнала наподобявайки възприятието на човешкия слухов апарат.
- В експериментите два от методите показаха предимство пред останалите. Тези са методът с много-лентово спектрално изваждане и методът с оценяване на кратковременната спектрална амплитуда, които използва оценител основаващ се на минимизация на средно квадратичната грешка на логаритмичния спектър.

Изследванията са финансирани по проекта MoveOn (IST-2005-034753) "Multi-modal and multi-sensor zero-distraction interaction interface for two wheel vehicles ON the move", съфинансиран по шеста рамкова програма (FP6) на Европейския Съюз.

В.УЧЕБНИ ПОСОБИЯ

[В1]. **Ганчев Т.**, В. Вълчев, Г. Николов, А. Маринов. *Материали и компоненти в електрониката*, Ръководство за лабораторни упражнения, ISBN:978-954-20-0578-0, Технически Университет - Варна, 2013.

Ръководството за лабораторни упражнения е съставено съгласно утвърдената учебна програма по дисциплината „Материали и компоненти в електрониката" (МКЕ), изучавана в първи курс от студентите със специалност „Компютърни системи и технологии" (КСТ). Подходящо е и за други специалности изучаващи базова електроника. Ръководството включва петнадесет теми за лабораторни упражнения, въвеждащи студентите в основни понятия от областта на електрониката и позволяващи придобиването на базови знания за дискретните полупроводникови прибори, съставляващи елементната база на съвременната електроника. Разглеждат се основните свойства и се изследват базовите характеристики на широк диапазон от дискретни електронни компоненти. Специално внимание е отделено за запознаване с полупроводникови прибори използващи новите силициево-карбидни и галиево-нитридни материали. Упражненията са организирани както следва:

Тема 1: Встъпително упражнение, запознава студентите с измервателната апаратура и правилата за безопасна работа в лабораторията.

Тема 2: Изследване на сензори за измерване на електрични величини, разглежда основните методи за измерване на електричните величини ток и напрежение и най-често използваните за целта сензори.

Тема 3: Изследване и измерване на основните параметри на пасивни елементи, запознава студентите със свойствата на градивните елементи резистор и кондензатор.

Тема 4: Изследване на изправителни диоди, е посветена на изследване на статичните и честотни характеристики на изправителните диоди.

Тема 5: Изследване на диоди със специално предназначение, изследват се някои диоди със специално предназначение, като светодиоди и ценерови диоди.

Тема 6: Изследване на биполярни транзистори, изследват се характеристиките на биполярен транзистор в схема на свързване общ емитер.

Тема 7: Изследване на статични характеристики на биполярни транзистори, изследват се характеристиките в схеми обща база и общ колектор.

Тема 8: Изследване на статични характеристики на полеви транзистори, изследват се статичните характеристики на полеви транзистор с PN-преход.

Тема 9: Изследване на полупроводникови елементи в ключов режим.

Тема 10: Изследване процесите на комутация при полевите транзистори и тяхното значение за компютърната техника.

Тема 11: Изследване на полупроводникови елементи базирани на съвременни материали и технологии, изследват се съвременни елементи като силициево-карбидни диоди и транзистори.

Теми 12: Изследване на оптоелектронни елемент и фотоприемници.

Тема 13: Изследване на светодиодни и течно-кристални индикатори.

Тема 14: Изследване на сензори за измерване на неелектрични величини.

Тема 15: Реализация на печатни платки.

Приложения: Пет приложения предлагат допълнителна информация на студентите, относно правилата за безопасна работа в лабораторията, предавателните характеристики на изследваните сензори изискванията към протоколите от лабораторни упражнения и точковата система за оценяване.

[В2]. Колев Й., Ц. Стоянова, В. Пенчев, И. Булиев, **Т. Ганчев**. *Микропроцесорни системи*. Ръководство за лабораторни упражнения, ISBN 954-8284-98-7, Технически Университет-Варна, 2000.

Ръководството за лабораторни упражнения е съставено съгласно утвърдената учебна програма по дисциплината „Микропроцесорни системи", изучавана от студентите от специалност „Електроника" на ТУ-Варна.

Ръководството за лабораторни упражнения включва четиринадесет теми, въвеждащи студентите в основни понятия от областта на микропроцесорната техника и позволяващи придобиването на базови знания за схемотехниката и програмирането на микропроцесорни системи. Цялостно се разглежда фамилията СМ600 от интегралните схеми българско производство, съставляваща елементната база за изграждане на индустриални контролери и системи за събиране на данни.

По-голямата част от лабораторните упражнения използват промишлените микропроцесорни системи „ИЗОМАТИК 1001УК" изградена на базата на осем битов микропроцесор СМ601. Тази промишлена система предлага класически решения, и е удобно средство за изграждане на микропроцесорни системи за събиране на информация и системи за управление на процеси и обекти.

Темите за лабораторните упражнения са както следва:

Тема 1: Създаване и настройка на програма на език асемблер за микропроцесор СМ601.

Тема 2: Универсален контролер „ИЗОМАТИК 1001УК". Структура, модули, програмно осигуряване – Мониторна програма „UKBUG". Работа с терминална програма „XTALK".

Тема 3: Предаване на данни по асинхронен интерфейс RS232C. Асинхронен сериен интерфейсен адаптер (АСИА) СМ603.

Тема 4: Прекъсвания от типа IRQ. Програмируем таймерен модул I8253.

Тема 5: Прекъсвания от типа NMI. Защита на данните в паметта на "ИЗОМАТИК 1001УК" при отпадане на захранването.

Тема 6: Схема за диагностика и индикация на грешка в "ИЗОМАТИК 1001УК".

Тема 7: Модул релейни изходи М240 за ИЗОМАТИК 1001УК.

Тема 8: Модул цифрови изходи К210 за ИЗОМАТИК 1001УК.

Тема 9: Модул цифрови входове К115 за ИЗОМАТИК 1001УК.

Тема 10: Модул аналогови входове К121 за ИЗОМАТИК 1001УК.

Тема 11: Модул аналогови изходи К220 за ИЗОМАТИК 1001УК.

Теми 12: Подготовка на приложни програми за запис в EPROM.

Тема 13: Работа с логически анализатор РМ3655.

Тема 14: Управление на клавиатура и индикация чрез ресурсите на ИЗОМАТИК 1001УК.

Приложения: Описание на логически анализатор РМ3655.

Изданието е частично финансирано по проекта Tempus „S-JEP 12456-97".

Дата: 23.08.2013

Изготвил:

/ас. д-р инж. Тодор Ганчев/