

РАЗПОЗНАВАНЕ НА ЗВУК ЧРЕЗ КОРЕЛАЦИОНЕН АНАЛИЗ

Иван Гарванов, Пенка Пергелова

Университет по библиотекознание и информационни технологии

Резюме: В настоящата статия е извършен корелационен анализ за разпознаване на звуци, генерирани от автомат „Калашников“. Анализирани са различни звукови характеристики, като е установено, че мелчестотните кепстрални коефициенти са най-ефективни за разпознаването на тези звуци. Този метод за разпознаване може да бъде използван и за идентифициране на различни източници на звукови сигнали, включително такива, генерирани от хора, птици, предмети, природни явления и др.

Ключови думи: разпознаване на звук, звукови характеристики, корелационен анализ.

Въведение

Задачите за разпознаване на звук са фундаментални за теорията на акустиката. През последните десетилетия се предлагат различни решения и приложения на звука в различни физични среди, като: подводна акустика, въздушна акустика, акустика в твърди тела [1]. Една от най-често решаваните задачи в практиката е гласово разпознаване на базата на техники за извличане на звукови характеристики. Обичайно използваните характеристики са: Кепстралният коефициент на перцептивното линейно предсказване (Perceptual Linear Prediction) (PLP), Коефициент за линейно предсказване (Linear Predictive Coding) (LPC), Мелчестотни кепстрални коефициенти (Mel-Frequency Cepstral Co-efficient) (MFCC), Методологии за автоматична идентификация на говорещи чрез динамично изкривяване на времето (Dynamic Time Warping) (DTW), Относителен спектрален анализ (Relative Spectral Algorithm) (RASTA), Метод на преминаване през нулата с пикови амплитуди (Zero Crossings with Peak Amplitudes) (ZCPA) и много други [2, 3, 4].

В [5] са изследвани акустични характеристики, извлечени от времевата и честотната област на звуков сигнал с цел откриване и разпознаване на безпилотни летателни апарати. От спецификата на звуковия сигнал произтича и изборът на изследваните аудиохарактеристики. Тези характеристики може да се използват в методите за машинно и/или дълбоко обучение в изкуствена невронна мрежа за откриване на безпилотни летателни апарати.

В [6] е изследван метод за разпознаване на звук от огнестрелно оръжие чрез спектрален анализ. Идентифицирането на вида оръжие по звука от изстрел намира приложение във военното дело, при реални бойни действия и подпомага балистичните експертизи при наличието на аудиозаписи от изстрели.

Акустичните характеристики, използвани в тази статия за анализ на звукови сигнали, са: спектър, спектрограма на сигнала, мелчестотни кепстрални коефициенти, гаматон кепстрални коефициенти, спектрален центроид, спектрално отклонение, пресичане през нулата и спектрална ентропия. Разпознаването на звуковите сигнали е реализирано с помощта на корелационен анализ на мелчестотни кепстрални коефициенти. Предложеният алгоритъм е тестван със звук от огнестрелно оръжие – автомат „Калашников“.

Звукови характеристики

Спектър на сигнала

Спектърът на сигнала се получава чрез преобразование на Фурие, което преобразува сигнала от времевата в честотната област чрез следния израз:

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt. \quad (1)$$

За да се приложи преобразованието на Фурие, трябва да са изпълнени условията на Дирихле и сигналът да е абсолютно интегрируем, т.е. интегралът на неговия модул да е краен:

$$\int_{-\infty}^{\infty} |x(t)| dt < \infty. \quad (2)$$

Модулът на спектралната функция се нарича амплитуден спектър, а неговият аргумент – фазов спектър. Преобразованието на Фурие свързва времевия сигнал със спектралната му функция, запазвайки цялата информация. Представянето на сигнала в честотната област съдържа същото количество информация, както и оригиналният сигнал във времевата област [6].

Спектрограма на сигнала

Съвкупността от моментните спектри на даден сигнал във функция на времето представлява спектрограмата на сигнала:

$$F(\omega, t) = \int_{t-T}^t X(t)e^{-j\omega t} dt. \quad (3)$$

За да се получи спектрограмата, е необходимо да се раздели времеви-ят сигнал на сегменти и да се приложи преобразованието на Фурие за всеки от тях [7]. Наборът от спектри образува спектрограмата на сигнала. Ако

сегментите на анализирания сигнал не се припокриват, то разделителната способност по честота се определя от зависимости $\Delta f = \frac{1}{T_1}$, а разделителната способност по време – от стойността на T_1 . Ако обработваните сегменти се припокриват, тогава разделителната способност по време е равна на $\Delta t = \frac{T_1}{N_1}$, където N_1 е броят на извадките от сегмента, подложен на преобразованието на Фурие (следователно числото N_1 се нарича размер на преобразованието на Фурие и е кратно на степен две). При висока степен на припокриване изчисленията могат да станат неприемливо много. Тъй като звуковият сигнал, генериран при изстрел, е с по-голяма продължителност, както и поради различните му честотни характеристики във времето, е поуместно да се анализира неговата спектрална диаграма [6]

Мелчестотен кепстрален коефициент (MFCC)

Мелчестотните кепстрални коефициенти (MFCC) са едни от най-популярните методи при извличане на спектрални характеристики. Освен че първоначално са получени за изследване на ехото в сеизмичния звук, те се използват и за моделиране на характеристиките на човешкия глас [8], както и в системите за разпознаване на реч.

Анализът на MFCC включва прилагане на бързо преобразование на Фурие (FFT) върху последователност от кадри, за да се получат определени параметри. Тези параметри преобразуват спектъра на мощност в спектър на мелчестота, след което се взема логаритъм и се извършва обратно преобразование на Фурие, както е показано на фиг. 1 [4].

Мелчестотните кепстрални коефициенти описват цялостната форма на спектралната обвивка:

$$M_n = \sum_{k=1}^k \log(S_k) \cos \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{k} \right], \quad (4)$$

където: S_k е изходът на филтърната банка, а k е броят на мел кепстралните коефициенти [5].



Фиг. 1. Анализ на мелчестотен кепстрален коефициент (MFCC)

Гаматон кепстралните коефициенти (GTCC)

Гаматон кепстралните коефициенти моделират човешката слухова реакция и спектралния анализ в кохлеята [9]. Някои автори предпочитат тези коефициенти за класификационни задачи. Въпреки че главно се използват за разпознаване на реч, понякога се използват и за откриване на аудио без говор [10, 11]:

$$G_n = \sqrt{\frac{2}{k}} \sum_{n=1}^k \log(S_k) \cos \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{k} \right] \quad (5)$$

където: S_k е изходът на банката филтърни гама тонове, а k е броят на гаматон филтърните коефициенти [5].

Спектрален центроид

Спектралният центроид показва върху коя честота е центрирана енергията на спектъра. Това е подобно на средно претеглена стойност:

$$SC = \frac{\sum_f D_x(f) \cdot f}{\sum_f D_x(f)}, \quad (6)$$

където: f е честотният индекс, а D_x е спектър на мощността на входния сигнал x [5, 12].

Спектрално отклонение

Спектрално отклонение (spectral rolloff point – SROP) е граничната честота, под която са концентрирани 95% от мощността на преобразованието на Фурие.¹

ZCR-скорост на преминаване през нулата

Скоростта на преминаване през нулата определя колко пъти сигналът преминава през нулевата стойност. Този параметър дава косвена информация за основната честота [5].

Спектрална ентропия

Спектралната ентропия (Spectral entropy) (SE) на сигнала е нелинеен параметър, който представя спектралното разпределение на неговата мощност. Тя е базирана на ентропията на Шанън, или информационната ентропия в теорията на информацията [13].

За да се изчисли ентропията за разпределение на кратковременен спектър на говорен сигнал, той се изследва като вероятно разпределение на дискретна случайна величина.²

Методология на изследването

В статията се използва корелационен анализ на мелчестотни кепстрални коефициенти за разпознаване на звукови сигнали, тъй като те са често използвани в практиката. Алгоритъмът е тестван с реални сигнали в средата на MATLAB.

Чрез корелационния анализ се изследва взаимовръзката между променливите X и Y , като се определят корелационните коефициенти и зависимостта между тях. Основни величини в анализа са корелацията и корелационното отношение, като търсената зависимост се изразява с числова оценка – коефициент на корелация. Степента на корелираност се определя от стойностите на коефициента на корелация, като по-близката до единица стойност указва по-силна зависимост между компонентите X и Y .

Корелационната зависимост се оценява с корелационното отношение R , което приема стойности в интервала $[0,1]$. Когато R е 0, няма корелационна връзка между Y и X . При $R = 1$ има функционална зависимост, докато при $0 < R < 1$ зависимостта е стохастична.³

Двете променливи си съответстват, когато зависимостта между тях е идеална, при $R = 1$ или $R = -1$, като посоката на зависимост между тях се определя от знака пред R . При положителна стойност на R двете променливи X и Y нарастват пропорционално. Когато R е с отрицателна стойност

при нарастване на X , променливата Y намалява. Изчислява се по формулата:

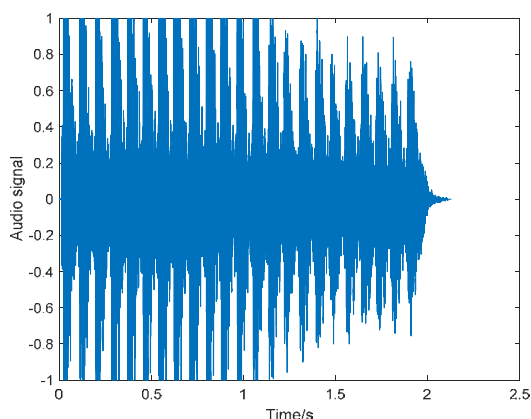
$$R_{xy} = \frac{cov_{XY}}{\sigma_X \sigma_Y} = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2 \sum(Y-\bar{Y})^2}} \quad (7)$$

където: $\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t$ и $\bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t$ са средни стойности на величините от n на брой измервания [14, 15].

Резултати

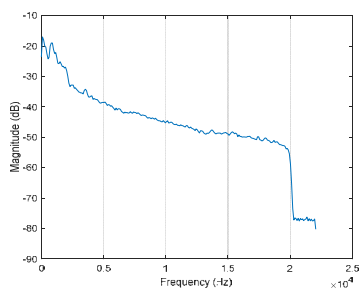
За разпознаване на звук от автомат „Калашников“ е проведено изследване на звуковите характеристики на генерираните от оръжието звуци. За определяне на тяхната релация е приложен корелационен анализ. Използваният в изследването звук е с продължителност от около две секунди и е показан на фиг. 2. Визуализира се амплитудата на звука във функция на времето, като за времето на изследване са произведени 23 изстрела.

Целта на настоящото изследване е да се определи корелационната зависимост между отделните изстрели.

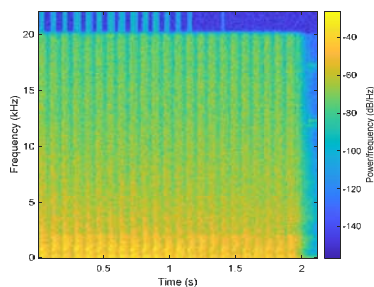


Фиг. 2. Звуков сигнал във функция на времето от автомат „Калашников“

След прилагане на бързо преобразование на Фурие към звука от фиг. 2, е получен спектърът на получения сигнал, който е показан на фиг. 3. Този спектър представя информация за честотата на сигнала за целия времеви диапазон на звука. По-детайлна представа за спектъра на сигнала във функция на времето би предоставила време-честотната диаграма (спектрограма), показана на фиг. 4.

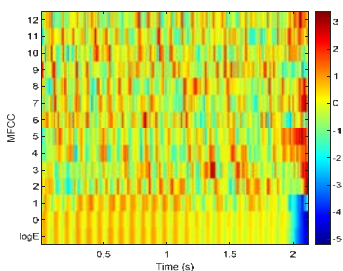


Фиг. 3. Спектър на звуковия сигнал, получен от автомат „Калашников“

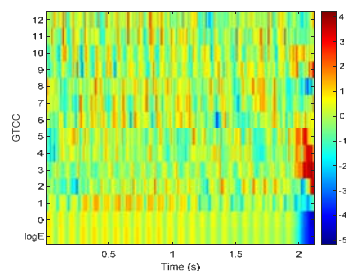


Фиг. 4. Спектрални диаграми на звуковия сигнал от автомат „Калашников“

При анализа на звуковия сигнал са получени част от звуковите характеристики на сигнала, като всяка от тях разполага със специфични черти, които могат да помогнат за разпознаването на звуковите сигнали. Мелчестотните кепстрални коефициенти и гаматон кепстралните коефициенти на изследвания звуков сигнал са показани на фиг. 5 и 6.

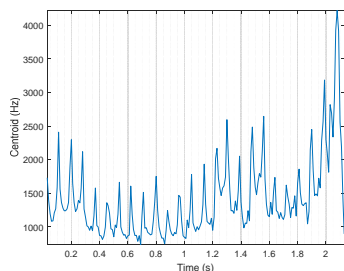


Фиг. 5. Мелчестотни кепстрални коефициенти (Mel Frequency Cepstral Coefficients – MFCC)

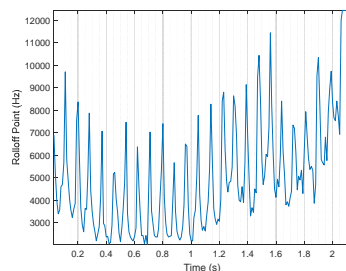


Фиг. 6. Гаматон кепстрални коефициенти (Gammatone cepstral coefficients – GTCC)

Спектрален центроид (Centroid) и спектралното отклонение са показани на фиг. 7 и 8.

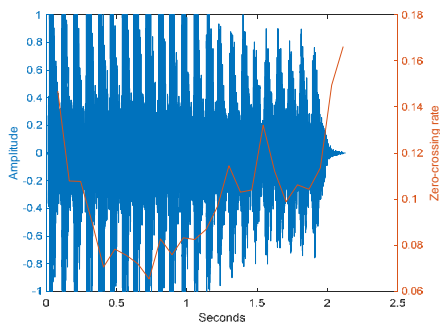


Фиг. 7. Спектрален центроид (Centroid)

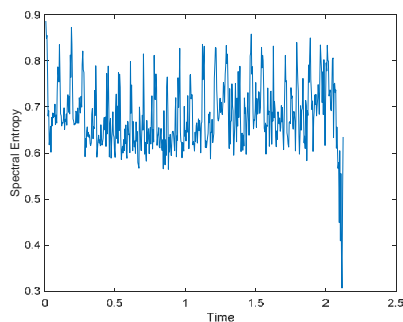


Фиг. 8. Спектрално отклонение

Характеристиките пресичане през нулата и спектрална ентропия са показани на фиг. 9 и 10.



Фиг. 9. Пресичане през нулата



Фиг. 10. Спектрална ентропия

След анализ на получените резултати и извършен литературен преглед бе установено, че мелчестотните кепстрални коефициенти са най-подходящи за разпознаване на звукови сигнали. На базата на това бяха изчислени корелационни коефициенти за някои от изстрелите в записа, като бяха анализирани изстрел 1, 5, 10, 15 и 20. Тези корелационни коефициенти са представени в таблица 1.

Таблица 1. Корелационни коефициенти

	Изстрел 1	Изстрел 5	Изстрел 10	Изстрел 15	Изстрел 20
Изстрел 1	1	0.87	0.89	0.93	0.89
Изстрел 5	-	1	0.93	0.89	0.85
Изстрел 10	-	-	1	0.91	0.93
Изстрел 15	-	-	-	1	0.95
Изстрел 20	-	-	-	-	1

Резултатите показват, че изстрелите от автомат „Калашников“ имат много сходни звукови характеристики, което ги прави лесни за разпознаване. Коефициентът на корелация между тях е около 0.9.

Заклучение

В настоящата статия са получени някои от най-популярните звукови характеристики, които могат да се използват при разпознаването на звукови сигнали. Посредством корелационен анализ са получени корелационни коефициенти на звуци, генерирани от автомат „Калашников“, които показват висока степен на близост и могат да се използват за разпознаване. Корелационният анализ е приложен на мелчестотни кепстрални коефициенти. Изс-

ледваният метод за разпознаване може да се използва и за решаване на други задачи, свързани с идентифициране на различни звукови сигнали, произведени от хора, птици, предмети, природни явления и други източници.

Благодарности: Тази статия е реализирана по Национална научна програма „Сигурност и отбрана“, финансирана от Министерството на образованието и науката в изпълнение на Националната стратегия за развитие на научните изследвания 2017-2030 и приета с Решение на Министерския съвет № 731 от 21 октомври 2021 г. Този материал отразява единствено мнението на авторите и МОН не носи отговорност за съдържанието.

Бележки

- ¹ **Pinquier, J.** Indexation sonore: recherche de composantes primaires pour une structuration audiovisuelle, Julien PINQUIER, Human-Computer Interaction. Universit e Paul Sabatier – Toulouse III, 2004. French. <tel-00008755> <https://theses.hal.science/tel-00008755> (31.05.2024).
- ² **Uzunov, A.** Detekcia na govor v sistemi za razpoznavane na diktori. Disertacia, 2020. <https://www.iict.bas.bg/konkursi/2020/AUzunov/disertacia.pdf> (31.05.2024).
[**Узунов, А.** Детекция на говор в системи за разпознаване на диктори. Дисертационен труд, 2020. <https://www.iict.bas.bg/konkursi/2020/AUzunov/disertacia.pdf> (31.05.2024).]
- ³ **Тема 13.** Statisticheski analiz na eksperimentalni dannii. – V: *Kurs ot lekicii na specialnost „Kompjuturni sistemi i tehnologii“*, TU Gabrovo. <<http://kst.tugab.bg/dl/week4/KMOD-Lab-4.pdf>> (17.05.2024).
[**Тема 13.** Статистически анализ на експериментални данни. – В: *Курс от лекции на специалност „Компютърни системи и технологии“*, ТУ Габрово. <<http://kst.tugab.bg/dl/week4/KMOD-Lab-4.pdf>> (17.05.2024).]

References/Литература

1. **Li, Y.** Entropy and Information Theory in coustics. – In: *Entropy* 2022, 24, pp. 1760, 2022. <https://doi.org/10.3390/e24121760>.
2. **Goyal, S., N. Batra.** Issues and Challenges of Voice Recognition in Pervasive Environment. – In: *Indian Journal of Science and Technology*, 2017, Vol. 10(30). DOI: 10.17485/ijst/2017/v10i30/115518, ISSN (Print): 0974-6846, ISSN (Online): 0974-5645.
3. **Kurzekar, P., R. Deshmukh, B. Vishal, P. Shrishrimal.** A Comparative study of Feature Extraction Techniques for Speech Recognition System. – In: *International Journal of Innovative Research, Engineering and Technology*, 2014, Dec. 3(12), pp. 1 – 11. DOI: 10.15680/IJRSET.2014.031203.
4. **Prithvi, P., T. Kishore.** Comparative Analysis of MFCC LFCC RASTA-LPP. – In: *International Journal of Scientific Engineering and Research*, 2016, 4(5), pp. 1 – 4.
5. **Garvanova, M., I. Garvanov.** Akustichni harakteristiki za otkrivane i razpoznavane na dronove. – V: *Proceedings of the XXI International Congress “Machines.*

Technologies. Materials” – Winter Session, 6 – 9 March 2024, Borovets, Bulgaria, 2024, Vol. 1, Issue 1 (28), pp. 98 – 101. ISSN: 2535-0021 (Print), ISSN: 2535-003X.

[**Гарванова, М., И. Гарванов.** Акустични характеристики за откриване и разпознаване на дроне. – В: *Proceedings of the XXI International Congress “Machines. Technologies. Materials”* – Winter Session, 6 – 9 March 2024, Borovets, Bulgaria, 2024, Vol. 1, Issue 1 (28), pp. 98 – 101. ISSN: 2535-0021 (Print), ISSN: 2535-003X.]

6. **Garvanov, I., P. Pergelova.** Razpoznavane na izstrel ot ognestrelno orazhie, izpolzvaiyki spektralnen analiz. – V: *Obshtestvoto na znaniето i humanizmat na XXI vek. XXI natsionalna nauchna konferentsia s mezhdunarodno uchastie*, Sofia, 1 noemvri 2023 g. Sofia: Za bukвите – O pismeneh, 2023, s. 336 – 346. ISSN: 2683-0094.

[**Гарванов, И., П. Пергелова.** Разпознаване на изстрел от огнестрелно оръжие, използвайки спектрален анализ. – В: *Обществото на знанието и хуманизмът на XXI век. XXI национална научна конференция с международно участие*, София, 1 ноември 2023 г. София: За буквите – О писменехъ, 2023, с. 336 – 346. ISSN: 2683-0094.]

7. **Garvanova, M.** Efekti ot prekomerna upotreba na smarttehnologii varhu choveka. Obrabotka na danni ot empirichni i eksperimentalni izsledvania. Monografiya. Sofia: Za bukвите – O pismeneh, 2022, 260 s. ISBN: 978-619-185-555-1, eISBN: 978-619-185-554-4.

[**Гарванова, М.** Ефекти от прекомерна употреба на смарттехнологии върху човека. Обработка на данни от емпирични и експериментални изследвания. Монография. София: За буквите – О писменехъ, 2022, 260 с. ISBN: 978-619-185-555-1, eISBN: 978-619-185-554-4.]

8. **Dwivedi, D., A. Ganguly, V. Haragopal.** Haragopal Contrast between simple and complex classification algorithms. – In: *Statistical Modeling in Machine Learning*, 2023. ISBN: 978-0-323-91776-6 с.
9. **Utebayeva, D., L. Iipbayeva, E. T. Matson.** Practical Study of Recurrent Neural Networks for Efficient, Real-Time Drone Sound Detection: A Review. – In: *Drones*, 2023, 7(1), p. 26. <https://doi.org/10.3390/drones7010026>.
10. **Garvanov, I., P. Pergelova, N. Nurdaulet.** Acoustic System for the Detection and Recognition of Drones. Telecommunications and Remote Sensing. ICTRS 2023. – In: *Communications in Computer and Information Science*, Vol. 1990. Springer, Cham, 2023. https://doi.org/10.1007/978-3-031-49263-1_8.
11. **Samaras, S., E. Diamantidou, D. Ataloglou, N. Sakellariou, A. Vafeiadis, V. Magouliantis, A. Lalas, A. Dimou, D. Zarpalas, K. Votis, et al.** Deep Learning on Multi Sensor Data for Counter UAV Applications – A Systematic Review. – In: *Sensors*, 2019, 19 (22), p. 4837. <https://doi.org/10.3390/s19224837>.
12. **Garvanova, M., I. Garvanov.** Akustichna sistema za откриване na bezpilotni letatelni aparati. – V: *VII International Scientific Conference on Security “CONFSEC 2023”* – 4 – 7 December 2023, Borovets, Bulgaria, Vol. 1 (9), pp. 104 – 107. Scientific Technical Union of Mechanical Engineering Industry-4.0, 2023. ISSN: Print 2603-2945, ISSN Online: 2603-2953.

[**Гарванова, М., И. Гарванов.** Акустична система за откриване на безпилотни летателни апарати. – В: *VII International Scientific Conference on*

- Security “CONFSEC 2023” – 4 – 7 December 2023, Borovets, Bulgaria, Vol. 1 (9), pp. 104 – 107. Scientific Technical Union of Mechanical Engineering Industry-4.0, 2023. ISSN: Print 2603-2945, ISSN Online: 2603-2953.]
13. **Galvão, F., S. Alarcão, M. Fonseca.** Predicting Exact Valence and Arousal Values from EEG. – In: *Sensors* (Basel), 2021 May 14, 21(10), p. 3414. DOI: 10.3390/s21103414, PMID: 34068895, PMCID: PMC8155937.
14. **Gmurman, V.** Теория вероятностей и математическа статистика: учебное пособие для вузов. Москва: Высшая школа, 2004, 479 с. ISBN: 5-06-004214-6.
[Гмурман, В. Теория вероятностей и математическая статистика: Учебное пособие для вузов. Москва: Высшая школа, 2004, 479 с. ISBN: 5-06-004214-6.]
15. **Eliseeva, I., M. Uzbashev.** Obshtaia Teoria statistiki: Uchebnik. 4-e izdanie. Moskva: Finansi i statistika, 2002, 480 s. ISBN: 5-279-01956-9.
[Елисеева, И., М. Юзбашев. Общая теория статистики: Учебник. 4-е издание, переработанное и дополненное. Москва: Финансы и статистика, 2002, 480 с. ISBN: 5-279-01956-9.]

За авторите

Иван Гарванов е професор в катедра „Информационни системи и технологии“, УниБИТ. Неговите научноизследователски интереси са в областта на цифровите технологии, обработката на комуникационни, радарни и навигационни сигнали, информационните технологии.

За контакт с автора: i.garvanov@unibit.bg

Пенка Пергелова е докторант в катедра „Информационни системи и технологии“, докторска програма ИСТИКН, УниБИТ. Нейните научноизследователски интереси са в областта на разпространението на звуковите вълни и влиянието им върху човека

За контакт с автора: p.pergelova@unibit.bg

SOUND RECOGNITION THROUGH CORRELATION ANALYSIS

Ivan Garvanov, Penka Pergelova

University of Library Studies and Information Technologies

Abstract: In this paper, a correlation analysis is applied for the recognition of sounds generated by a Kalashnikov assault rifle. To conduct the research, several sound characteristics were analyzed and it was found that Mel frequency cepstral coefficients are the most suitable for sound recognition. The researched recognition method can also be applied in solving many other tasks related to the identification of different sources of sound signals generated by people, birds, objects, natural phenomena, etc.

Keywords: sound recognition, sound characteristics, correlation analysis.

About the Authors

Ivan Garvanov is a professor in the Department of Information Systems and Technologies, at ULSIT. His research interests are in the field of digital technologies, communication, radar and navigation signal processing, information technology.

To contact the Author: i.garvanov@unibit.bg

Penka Pergelova is a doctoral student in the department of “Information Systems and Technologies”, PhD program ISTICN at ULSIT. Her research interests are in the field of propagation of sound waves, their influence on humans.

To contact the Author: p.pergelova@unibit.bg