

ТЕХНИЧЕСКИ УНИВЕРСИТЕТ – ВАРНА

Пролет Стойчева Денева

ИЗСЛЕДВАНИЯ ПО АВТОМАТИЧНО РАЗПОЗНАВАНЕ НА БИОЛОГИЧНИ ВИДОВЕ ОТ ТЕХНИТЕ АКУСТИЧНИ ЕМИСИИ

А В Т О Р Е Ф Е Р А Т

на дисертация за получаване на образователната и
научна степен „ДОКТОР”

по докторска програма: „Електронизация“

към професионално направление : 5.2 „Електротехника, електроника и
автоматика“

Научен ръководител: проф. д-р инж. Тодор Димитров Ганчев

Рецензенти:

- 1.
- 2.

Варна, 2022 г.

Дисертационният труд е обсъден на.....в катедра „ЕТМ“ на катедрен съвет,
съгласно заповед на Ректора на ТУ-Варна № /..... г. и насочен за защита.

Автор: Пролет Стойчева Денева

Заглавие: Изследвания по автоматично разпознаване на биологични видове от техните акустични емисии

ТЕХНИЧЕСКИ УНИВЕРСИТЕТ – ВАРНА

Пролет Стойчева Денева

**ИЗСЛЕДВАНИЯ ПО АВТОМАТИЧНО РАЗПОЗНАВАНЕ
НА БИОЛОГИЧНИ ВИДОВЕ ОТ ТЕХНИТЕ АКУСТИЧНИ
ЕМИСИИ**

А В Т О Р Е Ф Е Р А Т

**на дисертация за получаване на образователната и
научна степен „ДОКТОР”**

Варна, 2022 г.

Дисертационният труд съдържа 124 страници, включително 22 фигури, 7 таблици, 19 групи формули, оформени в 4 глави, приноси на дисертационния труд, списък с публикациите на автора по темата на дисертационния труд и списък на използваната литература от 177 заглавия на латиница.

Защитата на дисертационния труд ще се състои на г. от ч. в на открито заседание на жури сформирано със заповед на Ректора №/..... г.

Материалите по защитата (дисертацията, рецензиите и становищата) са на разположение на интересуващите се в Докторантски център, стая 318 НУК.

ХАРАКТЕРИСТИКА НА ДИСЕРТАЦИОННИЯ ТРУД

1. Актуалност на проблема

Опазването на биологичното разнообразие на флората и фауната е от съществено значение за оцеляването и развитието на екосистемите в околната среда. Усилията вложени в това начинание несъмнено водят до положителни ефекти върху всички човешки дейности и до подобряване на качеството на живот. Глобалните инициативи за мониторинг на биоразнообразието обикновено изискват наблюдение на различните животински видове, обитаващи големи територии, за да се оценят практическите ефекти от определени мерки и действия прилагани в подкрепа на тяхното опазване.

Реализацията на мащабно наблюдение на биоразнообразието за големи региони е скъпо и трудноемо, тъй като изисква денонощно присъствие на висококвалифицирани експерти за продължителен период от време, в различни метеорологични и климатични условия. Естествените местообитания на животинските видове са неподходящи или опасни за човешко присъствие. Наличието на хора променя поведението на животните и често води до нежелателни последици и за двете страни. Дори частична автоматизация на действията би имала огромно значение за разширяването на обсега на наблюдение на животинските видове.

Автоматизираното извличане на информация от звукозаписи, събрани за продължителни периоди от време, е изключително желателно, тъй като ще позволи да се подобри обхватът на мониторинга на биологичното разнообразие и да се направи оценка на въздействието върху околната среда на определени човешки дейности. Автоматичното извличане на информация също улеснява разработването на технологични инструменти, които биха улеснили интерпретацията на важни акустични събития или общите тенденции на динамиката на популацията и баланса между видовете.

2. Цел и задачи на изследването

Основната цел на настоящата дисертация е разработване на методи за моделиране и автоматично разпознаване на акустични емисии от птици, бозайници, влечуги и насекоми в техните естествени местообитания.

Реализацията на основната цел изисква поэтапна теоретична разработка и практическо изпълнение на следните задачи:

- Изследване възможността за определяне на характеристичната мелодия чрез интерполация на спектралните пикове на основния хармоник на сигнала, определени за кратки времеви интервали.
- Избор на метод за параметризиране на характеристичната мелодия на птичи вокализации, за да се интерполират отделните сегменти от характеристичната крива, изчислена за всяко акустично събитие.
- Автоматизирано разпознаване на типовете срички на птичите вокализации.

3. Обект и предмет на изследване

Обект на настоящата дисертация е изследването на методи и средства за анализ и обработка на птичи вокализации, **с цел разпознаване на биологични видове и типови акустични събития.**

Предмет на изследване в дисертационния труд е оценка на приложимостта на съществуващи и създаване на нови методи и средства за филтриране, шумопотискане, сегментиране, параметризиране на акустични сигнали на птици, с цел тяхното автоматизирано разпознаване.

4. Методи на изследване

Дисертационният труд се базира на последователното и поетапно приложение на различни методи и алгоритми за целенасочена обработка на данни както следва:

- Предварителна обработка на биоакустични сигнали посредством методите на филтриране, сегментиране и конволюция с прозоречна функция.
- Методи за параметризация на аудио сигнал основаващи се на съпоставянето на времевата и честотната информация за локализация на енергията на сигнала (аудио спектрограма).
- Методи за определяне на характеристичната мелодия на птичи вокализации.
- Методи за автоматизирано извличане и параметризиране на характеристичната крива на акустични събития на птичи вокализации.
- Използване на алгоритмите и методите на машинното обучение за автоматизирано разпознаване на типовете срички на птичите вокализации.

5. Място на изследване

Изследванията и обработката на аудио сигналите са проведени в лабораторната база на катедра ЕТМ при ТУ – Варна.

6. Научна новост на изследването

Създаден и валидиран е метод за определяне траекторията на доминиращия честотен компонент, определящ характеристичната мелодия на птичи вокализации с ясно изразена хармонична съставна, регистрирани в дивата природа. Създаден и валидиран е метод за параметризиране на характеристичната мелодия на птичи вокализации, използващ смеси от Гаусови функции (Gaussian Mixture Models, GMM), за да интерполира отделните сегменти от характеристичната крива, изчислена за всяко акустично събитие. Създаден и валидиран е метод за автоматизирано разпознаване на типовете срички на птичите вокализации.

7. Практическа ценност на изследването

Практическата приложимост на получените от изследванията резултати се определя:

- Създадено е програмно осигуряване реализиращо предложените методи.
- Проведени са експериментални изследвания за определяне на оптималните параметри на моделите.
- Проведени са числови симулации за оценка работоспособността на предложените методи.

8. Аprobация на изследването

Основните етапи от разработването на теоретични и приложни резултати на дисертационния труд са докладвани и публикувани в следните научни форуми и издания:

Конференции:

- 1 доклад на XXI International Conference on Geometry, Integrability and Quantization, June 3-8, 2019, Varna, Bulgaria, **Scopus**;

- 1 доклад на International Conference on Biomedical Innovations and Applications – BIA, November 8-9,2019, Varna, Bulgaria, **Scopus**;

Списания:

- 3 статии в Списание Акустика, Съюз на електроника, електротехника и съобщения Година XVIII, Брой 18/2016, стр. 87-90; Година XIX Брой 19/2017, стр. 57-60; Година XX, Брой 20/2018, стр. 12-17;

- 2 статии в Списание КНТ при ТУ-Варна, Година XVI, Брой 1/2018, стр. 57-60, Година XIX, Брой 1/2021, стр. 71-77;

- „XXI International Conference on Geometry, Integrability and Quantization“ и „International Conference on Biomedical Innovations and Applications – BIA“ са индексирани в международната научна база от данни “SCOPUS”.

9. Публикации по дисертационния труд

Основните етапи от разработването на дисертационния труд са отразени в 7 публикации, списък на които е приложен в края на автореферата.

СЪДЪРЖАНИЕ НА ДИСЕРТАЦИОННИЯ ТРУД

ГЛАВА I. ЛИТЕРАТУРЕН ОБЗОР НА ИЗСЛЕДВАНАТА ПРОБЛЕМАТИКА

В Глава I са представени спецификите и предизвикателствата пред реализирането на автоматичното разпознаване на вокализации на различни биологични видове. Направен е обзор на наличните решения на проблематиката и на начините по които се прилагат постигнатите резултати. Анализирани и изследвани са възможностите за изпълнението на поставената задача с различните технологии за реализация.

Анализът на биоразнообразието на екосистемите в Природата е пряко свързан с автоматизираната обработка на биоакустичните сигнали и тяхното последващо прецизиране без намесата на човешкия фактор. Това би улеснило събирането и систематизирането на данните при провеждане на изследванията на видовете и би спомогнало за запазването на по-голям обем информация, която да бъде използвана при решаването на конкретна задача. За анализа и мониторинга на биоакустични сигнали се ползва структурна методика, която включва три основни процеса:

1.Предварителна обработка на биоакустичния сигнал.

2.Параметризиране.

3. Последваща обработка.

Тази широко възприета в научните среди методика работи добре за много задачи за биоакустичната класификация, включително и за аудио сигнали, които са записани на открито в Природата. В глава I е представен подробен литературен обзор на трите основни процеса на горепосочената методика. Обоснована е актуалността на проблема, необходимостта от прилагането на съвременни теоретични достижения и експериментални средства за обработка и класификация на биоакустични сигнали.

ГЛАВА II. МЕТОДИ ЗА ПРЕДВАРИТЕЛНА ОБРАБОТКА НА АУДИО ЗАПИСИ С ПТИЧИ ВОКАЛИЗАЦИИ

В Глава II се изследва възможността за определяне на характеристичната мелодия чрез интерполация през спектралните пикове на основния хармоник на сигнала, определени за кратки времеви интервали. Представено е концептуалното решение на проблема и е предложен метод за откриване траекторията на характеристичната мелодия в биоакустични записи на птици, регистрирани в дивата природа. Показани са числени резултати от прилагането на метода и са синтезирани изводи и конкретизирани на потенциалната приложимост на метода.

2.2 1-D методи за параметризация на аудио сигнали

При 1-D методите за аудио параметризиране, сигналът се разделя на поредица от сегменти с кратко времетраене, като всеки тях да представлява квазистационарна част от аудио сигнала. Предварителна обработка с цел елиминиране на флукуации и артефакти в сигнала и ограничаване на честотната лента.

1. Потискане на нежелани изменчивости, интерференции и фонов шум в аудио сигнала.
2. Изчисляване на описатели, преобразуване на описателите и подбор на представителен набор от дескриптори.
3. Изчисляване на производни дескриптори, които да допълват характеристичния вектор с динамична информация, т.е. дават информация за промяната на оригиналните дескриптори във времевата област.

2.3 Методи за параметризация, основаващи се на аудио спектрограмата

За да се получи 2-D представяне на аудио сигнали, се следват следните утвърдени стъпки за предварителна обработка на сигналите:

1. Промяна на честотата на дискретизация в зависимост от конкретната задача. Честотата на дискретизация, с която са регистрирани записите на акустични емисии, често са получени чрез различни цифрови устройства, поради което е необходимо записите да бъдат преобразувани и с уеднаквен честотен диапазон, подходящ за извличане на полезната информация при последваща обработка.
2. Прилагане на стратегия за намаляване на шума и премахване на честотните ленти, чиито диапазон не представлява интерес за конкретната задача.
3. Сегментиране на аудио записа на обособени акустични събития или сегменти с фиксирана продължителност.
4. Изчисляване на време-честотното разпределение на енергията на сигнала чрез спектрограмата.

2.4 Метод за определяне на характеристичната мелодия на птичи вокализации

Основните етапи при извличане на характеристичните мелодии от аудио записи на птици са:

1. Предварителна обработка на сигнала за потискане на шума и конкурентни акустични емисии.
2. Откриване на спектралните пикове за първия хармоник на сигнала.
3. Интерполация през откритите спектралните пикове.
4. Съставяне на система характеристични уравнения.

2.4.1. Предварителна обработка на сигнала

Предварителната обработка на сигнала включва следните етапи:

1. Намаляване на честотата на дискретизация на аудио сигнала, което спомага да се намалят изчисленията в последващите стъпки на обработката.
2. Високочестотно филтриране на дискретния сигнал, целящо да се намали влиянието на нискочестотните смущения и шумове от околната среда, дължащи се на вибрациите, породени от оборудването и движението на растенията /при наличие на вятър/ или шум, породен от други животински видове и смущения, които са извън интересуваният ни честотен диапазон.
3. Динамичният спектър на предварително обработения дискретен сигнал от времевата област се преобразува чрез дискретно преобразуване на Фурие, реализирано чрез последователно прилагане на бързо преобразование на Фурие за последователни сегменти от сигнала.

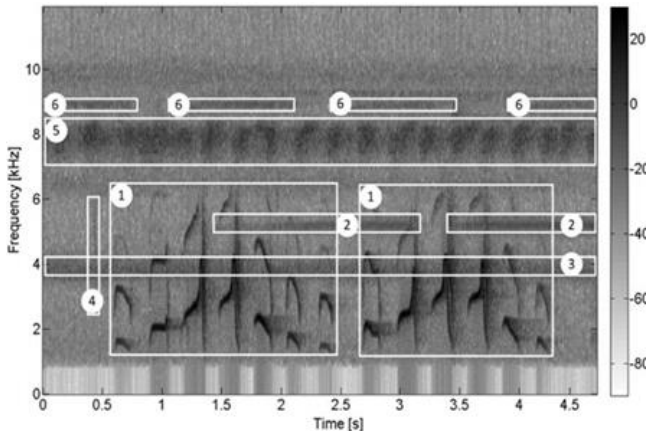
В настоящата разработка е използван 2-D метод за предварителна обработка на аудио сигнали, който е подходящ за шумопотискане в записи на птици. Този метод предлага добро потискане на фоновия шум в цялата честотна лента на сигнала и ефективно елиминира конкурентни акустични събития от повечето видове насекоми и земноводни. На Фиг. 2.2 е представен пример за необработен аудио запис направен в естественото местообитание на *Murmotherula multostriata*.

Спектрограмата от Фиг. 2.2 е изчислена за 5-секунден откъс от запис MYRMUL09.mp3, които е публично достъпен чрез базата данни Xeno-Canto . Звуците от Amazonian Streaked Antwren (*Murmotherula multostriata*) са отбелязани с правоъгълници с номера 1, а акустичните събития от други животински видове са отбелязани с правоъгълници с номера от „2“ до „6“. Звуковите събития с етикети / 2 / - / 6 / произхождат от други видове, които споделят едно и също местообитание. Като се има предвид, че се интересуваме от извличането на звукови дескриптори за звуковите събития, свързани с етикет / 1 /, трябва да се справим успешно с два ключови проблема:

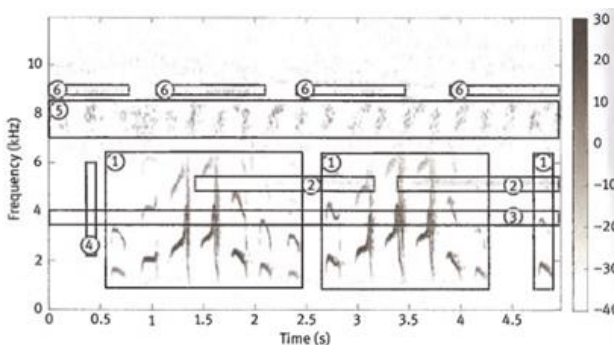
1. Има припокриване между звуковите събития на целевите видове / 1 / с тези на други биологични видове.
2. Някои звукови събития от други видове (напр. / 5 /) имат по - висока плътност на енергията от тази на целевите видове / 1 /.

Реалните записи на целевите видове обикновено съдържат шум от околната среда и звуци на множество други биологични видове. Качеството на суровата спектрограма не е достатъчно добро, за да извърши точен подбор на желаните високоенергийни области на целевите видове или да се изчислят надеждни описатели

на аудио сигнала. Един от методите за намаляване на влиянието на шума от околната среда в спектрограмата е да се нулират всички компоненти с амплитуди под определен праг както е показано на Фиг.2.3.



Фиг.2.2: Аудио спектрограма за откъс от пет секунди от запис на *Murmotherula multostriata* (акустични събития в правоъгълници с етикет 1), но включваща акустични събития от пет други животински вида (заградени в правоъгълници с етикети 2-6)[45].



Фиг.2.3:Аудио спектрограма след предварителна обработка на сигнала от Фиг.2.2 след прилагане на определен праг на амплитудата [45]

Морфологичното филтриране на спектрограмата е добре позната техника за изготвяне на изображения, която позволява селективно филтриране на изображението в зависимост от размера и формата на структуриращия елемент. Тъй като резултатът до голяма степен зависи както от съдържанието на изображението, така и от избора на размер и форма за структуриращия елемент, обикновено последният се дефинира въз основа на предварително познание за 2-D обектите, които бихме искали да премахнем, запазим, моделираме или разпознаем.

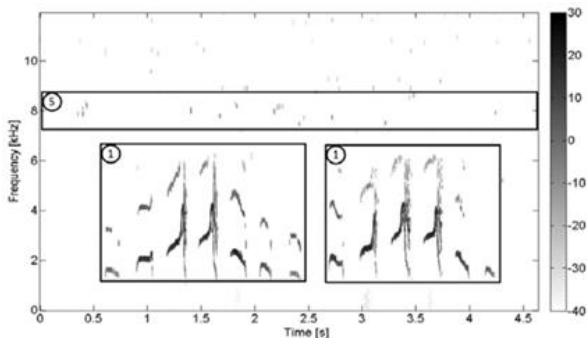
Морфологичните техники за обработка на изображения предоставят нов гъвкав начин за справяне с шума от околната среда и смущения, който допълва традиционните 1-D методи за обработка на сигнала.

На Фиг.2.4 е показан резултата от прилагането на метода за потискане на фонов шум и нежелани акустични събития, приложен върху спектрограмата от Фиг.2.2. В резултат на обработките, фоновия шум и съпътстващите акустични емисии от други животински видове, различни от Amazonian Streaked Antwren, са подтиснати с изключение на дребни фрагменти от акустичната емисия заградена с правоъгълник „5“.

Следващите стъпки на предварителна обработка на сигнала са елиминирането на остатъчния шум и възстановяването на оригиналните обекти в морфологично филтрираната спектрограма. Последваща обработка се състои от следните компоненти:

1. Оценка прага на решение, пропорционален на средната стойност на всички пиксели в морфологично филтрирана спектрограма.
2. Прилагане на прага върху морфологично филтрираната спектрограма, като предварително се запазят елементите с енергия над прага и се нулират тези под прага.
3. Замяна на стойността на запазените елементи с тези от оригиналната спектрограма.

За случаите на вокализации с изявени хармонични съставящи, има възможност да бъде направена апроксимация на характеристичната мелодия. За целта се започва с прилагане на маска за отделяне на първия хармоник, последвано от време-честотна локализация на пиковете на енергията.



Фиг.2.4: Аудио спектрограма след предварителна обработка на сигнала от Фиг.2.2 с 2-D метода [45]

Прилагането на маска и локализацията на пиковете на енергия в изображението от Фиг.2.4, дава възможност за определяне на малък набор от позиции които трасират изменението на максималната енергия по

време на вокализацията. Локализацията на позициите с пикова енергия във време-честотната област дава възможност за компактно представяне на вокализацията и за създаване на нови методи за разпознаване на голям брой животински видове по тяхната характеристичната мелодия.

2.4.2. Интерполация

Използвайки позициите с максимална концентрация на енергията, определени съгласно обработките на сигнала описани по-горе, с помощта на числова интерполация

можем да конструираме нови числови стойности, които да допълват липсващи участъци от характеристичната мелодия. Според конфигурацията и броя на получените позиции с максимална енергия може да се използва линейна, параболична, билинейна, бикубична или други видове функции за интерполиране. В конкретният случай приемаме параболичната интерполация като подходящ начин да намерим характеристичната мелодия.

Нека да приемем, че прилагането на обработката на сигнала е довело до определяне на координатите за всеки от спектралните пикове, представени в Таблица 2.1. При пресмятанята ще считаме, че стойностите y_i в точките на аргумента x_i са получени от експеримент и ние не познаваме стойностите на функцията y за стойности различни от x_i . За всяка комбинация от три последователни позиции се съставя параболично уравнение чрез интерполационен полином на Лагранж.

Таблица 2.1: Координати на спектралните пикове

x_i	x_0	x_1	x_2	x_3	$x_4 \dots$	x_m
y_i	y_0	y_1	y_2	y_3	$y_4 \dots$	y_m

В общият случай, получената система от уравнения се представя както следва:

$$\begin{aligned}
 Y_{0(x)} &= a_0 \cdot x^2 + b_0 \cdot x + c_0 \quad \text{за интервал } \Delta_0 = x_0 \div x_2 \\
 Y_{1(x)} &= a_1 \cdot x^2 + b_1 \cdot x + c_1 \quad \text{за } \Delta_1 = x_2 \div x_4 \\
 Y_{2(x)} &= a_2 \cdot x^2 + b_2 \cdot x + c_2 \quad \text{за } \Delta_2 = x_4 \div x_6 \\
 &\dots \\
 Y_{m-4} &= a_{m-4} \cdot x^2 + b_{m-4} \cdot x + c_{m-4} \quad \text{за } \Delta_{m-4} = x_{(m-2)} \div x_6
 \end{aligned}
 \tag{2.6}$$

Приемаме, че прилагането на предварителната обработката на сигнала е довело до определяне позициите с максимална енергия за най-значимите спектрални пикове от Фиг. 2.4, чиито примерни координати са представени в Таблица 2.2.

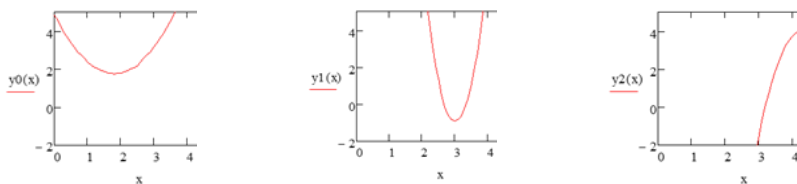
Таблица 2.2: Примерни позиции с максимална енергия

x_i	1.3	1.6	2.4	2.5	3.7	4.6	4.9
y_i	2	1.8	2.1	1.2	2.8	3.7	2.8

Прилагаме системата от уравнения (2.6) за всяка комбинация от три последователни позиции и съставяме параболичните уравнения.

Графичната интерполация на получената система уравнения, представена на Фиг. 2.5, създава визуална представа за елементите на характеристичната мелодия определени от уравненията. Въвежда се нов метод за описване на акустични емисии на животински видове, подходящ за сигнали, които имат ясно изразена хармонична съставна. Предложената методика създава предпоставки за създаване и развитие на нови методи за бързо търсене на определен тип вокализация в дълги записи, а също така, за

ускорено разпознаване на голям брой животински видове по тяхната характеристичната мелодия.



Фиг.2.5: Елементи на характеристична мелодия

2.4.3. Сегментиране на сигнала

Сигналят, обработен съгласно описаната по-горе методика, се подлага на сегментиране за отделяне на обособени акустични емисии. За целта бинаризираната траектория на пиковата енергия се подлага на конволюция с правоъгълна прозоречна функция с дължина 50 мс. Това позволява елиминиране на малки прекъсвания в траекторията на пиковата енергия, дължащи се на грешки при определянето ѝ или на наличието на съпътстващи акустични събития от други източници. В резултат на конволюцията участъците, в които липсват изявени вокализации се преобразуват в дълги последователности от нули. В настоящата разработка е прието акустичните събития, разделени с такива участъци от нули, да се приемат като отделни сегменти, ако нулите са в продължение на повече от 100 мс. Експериментално е установено, че такава граница позволява да се сегментират акустичните вокализации на множество птичи вокализации, в т.ч. и Amazonian Streaked Antwren. В резултат на приложения алгоритъм две акустични събития се обособяват като самостоятелни, ако отстоят на повече от 100 мс. Приложен към спектрограмата на Фиг.2.4, така описаният алгоритъм довежда до формиране на 12 обособени сегмента, които се апроксимират самостоятелно.

Използвайки позициите с максимална концентрация на енергията, определени съгласно предварителната обработката и сегментацията на сигнала, с помощта на числова интерполация се конструират нови числови стойности, които описват обособените участъци от характеристичната мелодия.

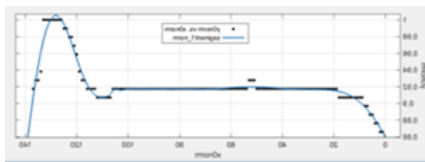
За целта се използва нормално (Гаусово) разпределение. Коефициентът a контролира ширината на камбаната, коефициентът b контролира централния максимум на камбаната по дължина на оста x , а коефициентът c контролира височината на камбаната. В настоящата разработка се използва модел с шест Гаусови функции:

$$f(x) = a_1 e^{-\left[\frac{(x-b_1)}{c_1}\right]^2} + a_2 e^{-\left[\frac{(x-b_2)}{c_2}\right]^2} + a_3 e^{-\left[\frac{(x-b_3)}{c_3}\right]^2} + a_4 e^{-\left[\frac{(x-b_4)}{c_4}\right]^2} + a_5 e^{-\left[\frac{(x-b_5)}{c_5}\right]^2} + a_6 e^{-\left[\frac{(x-b_6)}{c_6}\right]^2} \quad (2.8)$$

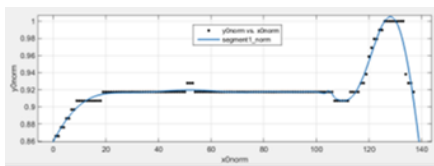
Показаният на Фиг.2.4 сигнал е сегментиран на две големи групи с общо 12 отделни сегмента. За три от тези, резултатите от моделирането с (2.8) са показани съответно на Фиг.2.6, Фиг.2.7 и Фиг.2.8.

Анализирайки времевите сегменти се получават следните коефициенти на апроксимиращите функции (Таблица 2.3).

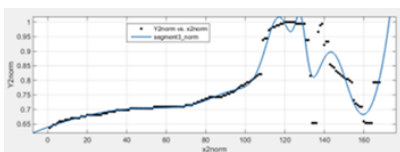
От получените експериментални резултати е очевидно, че при използване на характеристично описание със смес от Гаусови функции се получава добра апроксимация. Относително малката грешка между сигнала и апроксимиращата функция показва, че избраният модел позволява добро приближение до анализирания биоакустичен сигнал. След определяне на коефициентите в модела (2.8) се съставя системата от характеристични уравнения, описващи група сегменти от сигнала.



Фиг.2.6: Апроксимация на характеристичната мелодия за сегмент 1



Фиг.2.7: Апроксимация на характеристичната мелодия за сегмент 2



Фиг. 2.8: Апроксимация на характеристичната мелодия за сегмент 3

Такова компактно описание на акустичните събития създава условия за разработване на нови методи за параметризация на сигнала и за създаване на нови методи за разпознаване на голям брой животински видове по тяхната характеристичната мелодия.

Характеристичната мелодия на птичите вокализации е измежду най-информативните описатели при моделирането и автоматичното разпознаване на птици по техните акустични емисии. В настоящата дисертация се представя анализ на различни методи за апроксимиране на сегменти от време-честотното разпределение на енергията, получени от аудио спектрограмата на птичи вокализации. Изследвана е приложимостта на три методи за апроксимация: чрез Гаусови функции, чрез преобразуване на Фурие и чрез полиномиална функция и е направена оценка на абсолютната грешката в средно-квадратичен смисъл. Сравнението на методите е извършено по критериите: сумарна квадратична грешка (SSE) и квадратен корен от средно-

квадратичните грешки (RMSE). Експерименталните резултати показват, че методът за апроксимация чрез Гаусови функции води до най-малка абсолютна грешка и най-малка дисперсия на грешката, т.е. с най-добра предсказуемост на грешката.

В настоящата дисертация се изследва възможността за оптимално конструиране на характеристичната мелодия на птичи вокализации, чрез интерполация на основния хармоник на сигнала, определен за кратки времеви интервали, по различни методи за моделиране в среда от MATLAB.

За да бъдат анализирани и разпознати биоакустичните емисии на различните видове птици се прилага обработка на характеристичните мелодии от аудио записи. Извършва се интерполация на сигнала като се избира метод за интерполиране с най-малко отклонение от автентичния сигнал. Основните етапи при обработката на полезните акустични събития и тяхната оценка са:

1. Предварителна обработка на сигнала за потискане на конкурентни акустични емисии.
2. Разделяне на сигнала на кратки времеви интервали.
3. Интерполация на позициите с доминантна енергия на сигнала по различните методи в среда на MATLAB.
4. Оценка на грешката при моделиране на характеристична мелодия на птичи вокализации при интерполация на опитните резултати чрез Гаусови функции, чрез преобразуване на Фурие и чрез полиномиална функция.

Таблица 2.3: Изчислени стойности на параметрите на модела (2.8) за три обособени акустични събития от Фиг.2.4

C/S	Segment 1	Segment 2	Segment 3
a1	0,3562(-0,03164;0,7441)	27,06(11,81;42,31)	0,221(0,103;0,3391)
b1	1,561(-1,405;1,718)	1333(1332;1334)	128,1(127,3;129)
c1	0,4019(0,1842;0,6196)	3,275(1,497;5,053)	3,96(2,248;5,671)
a2	0,9161(0,8978;0,9344)	124,5(-193,6;442,5)	0,3478(-0,2405;0,9361)
b2	-0,3891(-1,535;0,7573)	1349(1289;1409)	143,6(140,6;146,6)
c2	3,203(-0,2408;6,647)	61,81(-48,24;171,9)	12,09(2,642;21,55)
a3	0,09361(-0,3748;0,562)	38,18(23,6;52,76)	0,3441(0,06392;0,6243)
b3	-1,66(-2,601;-0,7182)	1324(1322;1326)	118,1(115,9;120,4)
c3	0,65(-0,5028;1,803)	7,897(4,278;11,52)	10,28(4,922;15,64)
a4	0,00593(-0,004227;0,01609)	-2,512(-26,63;21,61)	1,188e+07(-4,849e+10;4,851e+10)
b4	0,9096(0,866;0,9531)	1286(1255;1316)	533,7(-8,574e+04;8,681e+04)
c4	0,03417(-0,03834;0,1067)	13,05(-56,5;82,6)	88,04(-1,027e+04;1,045e+04)
a5	0,07283(-0,2749;0,4205)	20,6(-367,9;409,1)	0,1198(-0,1475;0,3872)
b5	0,7045(-1,433;2,842)	1274(1127;1421)	101,4(63,85;139)
c5	0,5525(-1,029;2,134)	38,76(-160,8;238,3)	22,57(-16,05;61,19)
a6	0,04648(-0,326;0,419)	104,9(50,19;159,6)	0,7045(0,6817;0,7273)
b6	0,9264(0,7548;1,098)	1210(1024;1396)	50,96(8,389;93,53)
c6	0,2903(-0,2067;0,7874)	99,9(-333,4;533,2)	163(20,07;306)

Целта на апроксимацията е изчисляване на траекторията на характеристичната мелодия, посредством интерполация на позициите с доминантна енергия в аудио спектрограмата. В резултат на интерполацията се определят нови числови стойности, които описват обособените участъци от характеристичната мелодия съгласно предварително избран модел. Параметрите на модела се използват за дескриптори на конкретния сегмент от аудио спектрограмата.

Целта на настоящото изследване е определяне на пригодността на три различни метода за апроксимация, а именно: чрез смес от Гаусови функции; чрез преобразуване на Фурие; чрез полиномиална функция.

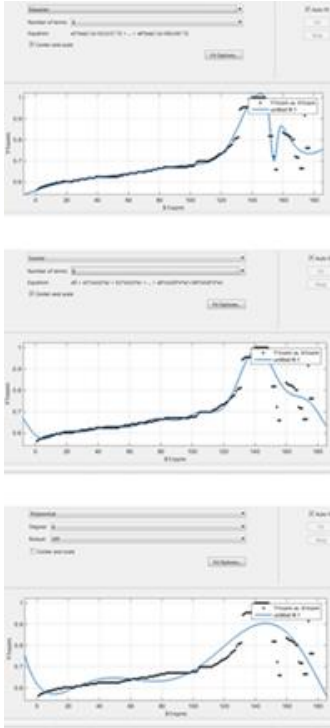
Сравнение на апроксимациите получени с тези три метода за първия сегмент от характеристичната мелодия от Фиг. 2.4, е представено на Фиг.2.9.

В Таблица 2.4 са представени стойностите на сумарната квадратична грешка (SSE) и квадратен корен от средно-квадратичните грешки (RMSE) получени за първите пет сегмента от общо дванадесетте открити за аудио спектрограмата от Фиг. 2.2.

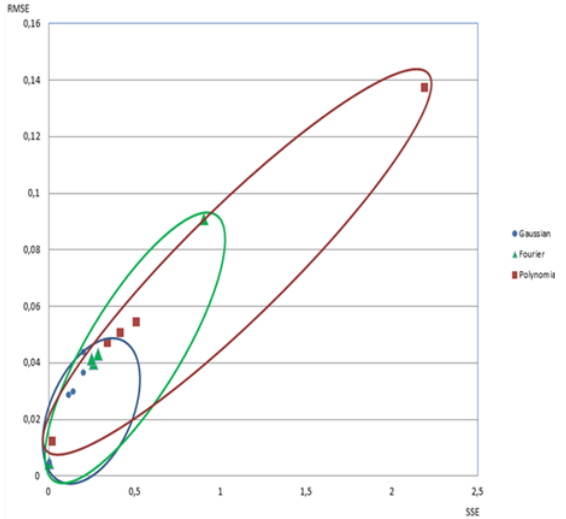
Таблица 2.4: Стойности на сумарната квадратична грешка (SSE) и квадратен корен от средно-квадратичните грешки (RMSE) получени за първите пет сегмента

Метод	Gaussian		Fourier		Polynomial	
	SSE	RMSE	SSE	RMSE	SSE	RMSE
I	0,0028	0,0049	0,0029	0,0049	0,0194	0,0122
II	0,1398	0,0297	0,2622	0,0401	0,5076	0,0546
III	0,2002	0,0365	0,2865	0,0431	0,4134	0,0507
IV	0,2008	0,0437	0,9016	0,0909	2,1864	0,1373
V	0,116	0,0287	0,2489	0,0414	0,3397	0,0473

Както е показано на Фиг. 2.9, методът за апроксимация чрез Гаусови функции води до най-малка абсолютна грешка и най-малка дисперсия на грешката, т.е. получените апроксимации са с най-добра предсказуемост на грешката.



Фиг.2.9: Апроксимация на сегмент чрез:(а) смес от Гаусови функции;(б) преобразуване на Фурие;(в) полиномиална функция.

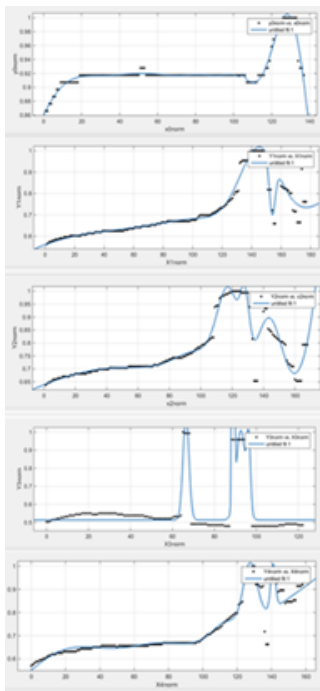


Фиг.2.10: Сравнение на трите метода за апроксимация по критерии SSE и RMSE

Сравнение на трите метода за апроксимация от Таблица 2.4, по критерии SSE и RMSE е представено на Фиг.2.10. По абсцисата е показана сумарната квадратична грешка (SSE), а по ордината квадратен корен от средно-квадратичните грешки (RMSE).

На Фиг.2.11 са представени резултатите от апроксимация на първите пет сегмента от спектрограмата, получени чрез най-успешния метод - смес от Гаусови функции със шест компонента. Както може да се отбележи, апроксимацията пресъздава достатъчно точно измененията на енергията за всеки един от сегментите.

От анализът на експерименталните резултати може да се потвърди направено предположение, че апроксимацията чрез смес от Гаусови функции е подходящ инструмент за моделирането на характеристичната мелодия на птичи вокализации. Оценката на средно квадратичната грешката показва, че при съизмерима сложност, методът използващ смес от Гаусови функции превъзхожда апроксимациите получени чрез преобразуване на Фурие и чрез полиномиалната функция.



Фиг.2.11: Апроксимация на характеристичната мелодия чрез Гаусови функции за първите пет сегмента от аудио спектрограмата.

Gaussian	I сегмент	II сегмент	III сегмент	IV сегмент	V сегмент
a1	0,3549	0,3587	0,221	0,5789	0,2535
b1	130,8	1,1	128,1	66,45	127,6
c1	15,93	0,3618	3,96	2,015	5,601
a2	0,9166	1,969	0,3478	0,513	0,1945
b2	55,2	6,268	143,6	96	140,9
c2	126,1	4,294	12,09	1,966e+05	1,944
a3	0,1034	-0,2542	0,3441	5,845	0,0608
b3	2,571	1,267	118,1	88,94	122,5
c3	26,75	0,05569	10,28	0,5714	10,64
a4	0,00603	0,124	1,188e+07	-5,578	1,703e+14
b4	105,1	0,388	533,7	88,96	1,121e+04
c4	1,354	0,9099	88,04	0,2742	1927
a5	0,04948	-0,00081	0,1198	0,4988	0,0827
b5	90,52	-0,4117	101,4	92,57	72,53
c5	17,6	0,01041	22,57	3,146	25,94
a6	0,07772	0,5131	0,7045	0,3858	0,2448
b6	105,6	-1,384	50,96	96,34	18,33
c6	12,83	2,123	163	1,225	40,89

Таблица 2.5: Параметри на модела със смес от шест Гаусови функции, получени при апроксимацията за петте сегмента от Фиг.2.4.

ИЗВОДИ към Глава II

Въз основа на изложението и експерименталната валидация, могат да се направят следните изводи:

Анализът на аудио спектрограмата се утвърждава като основен инструмент за изследване на биоакустични сигнали. Аудио спектрограмата позволява интуитивно

представяне на време-честотно разпределение на енергията на звуковия сигнал от гледна точка на човешкото възприятие, но тук е показана и нейната полезност при изчисляване на характеристикната мелодия на птичите вокализации.

Въз основа на информация извлечена от спектрограмата е предложен нов метод за описване на акустични емисии на животински видове, подходящ за сигнали, които имат ясно изразена хармонична съставна. Методът задава формална параметрична рамка, в която се дефинира изчисляването на характеристикна мелодия на акустични емисии на пойни птици.

Експериментално е проверено, че предложеното параметрично описание на характеристикната мелодия дава възможност за бързо сравнение между структурите и композициите на множество акустични емисии на пойни птици и други животински видове, които издават звуци с изявена хармонична структура.

ГЛАВА III. ИЗВЛИЧАНЕ НА ПАРАМЕТРИТЕ И ОПИСАТЕЛИТЕ НА ПТИЧИТЕ ВОКАЛИЗАЦИИ.

В Глава III се представя нов метод за параметризиране на характеристикната мелодия на птичи вокализации, който използва Gaussian Mixture Models (GMM), за да интерполира отделните сегменти от характеристикната крива, изчислена за всяко акустично събитие. За целта е предложен тристепенен процес на обработка на аудио сигнала, при който първоначално се извършва морфологична обработка и автоматично сегментиране на птичата песен за отделяне на акустични събития въз основа на аудио спектрограмата [45]. На втората стъпка се прилага алгоритъм за намиране на доминиращия честотен компонент, изглаждане на доминиращия енергиен контур и редуциране на изображението до черно и бяло. Предложен е метод за параметризация на полученния контур, т.е. характеристикната крива за всяко акустично събитие, чрез интерполация с GMM. Резултатът от този процес е параметричен модел, който описва сегменти с различна дължина и сложност с предварително определен брой параметри. За валидиране на приложимостта на предложения подход е представена експериментално изследване за параметризиране на акустични емисии на биологичният вид *Mugmotherula multostriata*, като за целта са използвани публично достъпни полеви записи регистрирани в естествените местообитания на тези птици. Експерименталните изследвания показват, че моделиране със шест-компонентен GMM осигурява необходимата прецизност и точност на представяне на характеристикната мелодия. Получените чрез този модел параметри могат да служат добре при моделиране на птичи вокализации или приложения за разпознаване на видове.

3.1 Въведение в проблема

Характеристичната мелодия на птичите песнопения добре описва цялостната динамика на вокализацията на птиците и основните параметри на звуковия източник, което го прави важно доказателство в казуси, свързани с разпознаването на видовете птици и проучвания върху репертоара на птичите песни.

Въпреки това птичите песни, записани в естествените местообитания на дивите видове и записите на звуковия пейзаж обикновено съдържат високи нива на шум от

околната среда и едновременни акустични емисии от други източници. Това прави акустичния фон променлив във времето и труден за моделиране и отделяне на вокализацията на целевите видове. В допълнение, често полевите записи съдържат едновременни звукови събития от множество животински видове и това е основната пречка за получаване на висока точност при автоматизираното акустично разпознаване на видове.

Разработките в Глава III надграждат върху предишни изследвания, които използват методи за намаляване на смущения, които действат едновременно на множество колони и редове на аудио спектрограмата. Тези методи използват предимството на съпадението на информацията за времето и честотата на изследвания обект и неговия математически модел и поради тази причина често се наричат 2-D методи за параметризиране на звука, като се прилага предварителната обработка, потискане на околния шум и сегментиране на записите на птиците, записани в естествените им местообитания.

Основната разлика по отношение на предишните разработки е, че в настоящата дисертация основният фокус е върху оценката на приложимостта на автоматизираната параметризация на характеристикната крива, с идеята да се получи автоматично сегментиране на птича песен и холистично представяне на аудио сегменти с променлива дължина и сложност. В този смисъл, за разлика от предишни разработки, тук фокусът не е да се изчислят 1-D дескриптори, базирани на отделен аудио кадър, като например Мел-кепстралните коефициенти (Mel-Frequency cepstral coefficient, MFCC) или някакъв друг тип двуизмерни дескриптори, които улавят свойствата на честотата във времето за малки сегменти от вокализацията на птиците или моделират разсейването на енергията. Вместо това основният фокус в настоящото изследване е върху оценката на приложимостта на автоматизираната параметризация на характеристикната крива, с идеята да се получи представяне на аудио сегменти с фиксирана дължина, които са с променлива сложност, получени след автоматично сегментиране на птича песен. Предполага се, че подобно представяне би се превърнало в удобна характеристика при автоматизираното разпознаване на видове птици и в технологично подкрепени изследвания върху акустичния репертоар на видове птици.

3.2 Метод за автоматизирана параметризация на акустични събития

На Фиг.3.1 е представена цялостната блок-схема на предложения метод за автоматизирана параметризация на акустични събития на птичи песни, която илюстрира двата основни етапа на обработка на звука, включително предварителната обработка и сегментиране на звука и параметризацията на акустичните събития. Етапът на предварителна обработка на звука изпълнява редица последващи стъпки във времевата и честотната област, които имат за цел да потиснат потенциалните акустични смущения от околната среда и от едновременни акустични събития от шумът на околната среда и нецелевите видове. Сегментацията на птичите песни е реализирана във времево-честотната област и има за цел да изолира отделните акустични срички. Параметризацията се основава на проследяване на доминиращата честотна компонента и автоматизирано сближаване на характеристикната крива на всяко акустично събитие. В крайна сметка характеристикната крива подлежи на интерполация с метод, основан на GMM. Резултатът от този процес е усъвършенстван модел на GMM, параметрите на

който могат да бъдат използвани като характеристични вектори в задачите за разпознаване на птичи песни и птичи видове.

3.3 Предварителна обработка на аудио сигнала

Аудио сигналът е необходимо да бъде предварително обработен за намаляване на шума и за разделяне на дълги аудио записи на парчета с продължителност от няколко секунди. В дисертационния труд е следван метода, описан в [45]. Първоначалните стъпки за намаляване на шума се състоят в ограничение на честотната лента, във времевата област, което има за цел да намали честотния диапазон до този, който има отношение към целевите видове. За тази цел, първо се прилага лентов филтър, целящ да намали различни смущения. Сред тях са нискочестотният шум поради вибрации на оборудването, директни удари на вятъра в микрофона, движения на растителността поради движение на въздушни маси и други нискочестотни смущения в околната среда, които се намират извън честотния диапазон, който ни интересува. Горната честота на сръза е избрана да бъде малко над най-високата честота, която ни интересува, но и все пак да елиминира високочестотните звукови емисии от прилепи и насекоми. Тогава лентово филтрираният сигнал се сегментира с прозореца на Хаминг, $W(m)$:

$$W(m) = 0.54 - 0.46 \cos\left(\frac{2\pi m}{M}\right), m = 0, 1, \dots, M - 1 \quad (3.1)$$

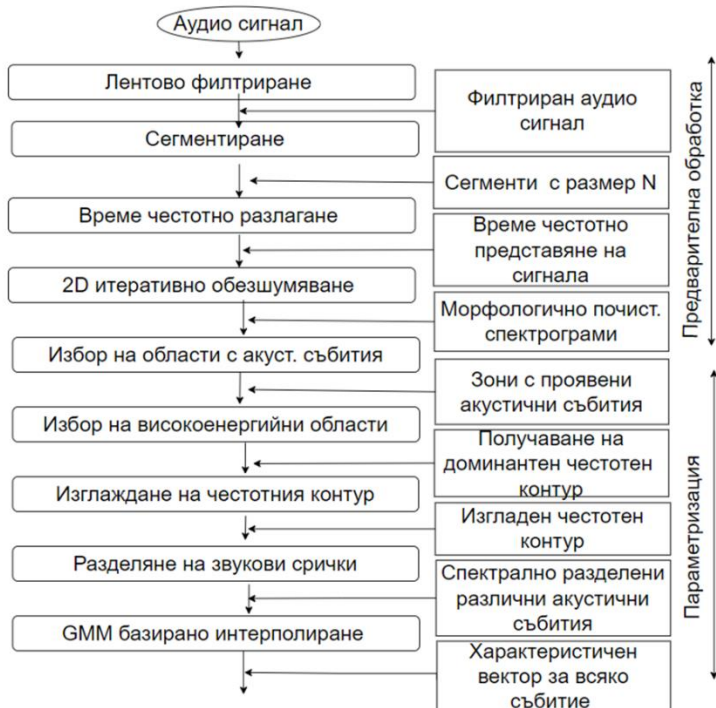
където дължината на прозореца е избрана така, че звуковият сигнал да се сръзва на кратки квазистационарни части с продължителност в диапазона на десетки милисекунди. След това се прилага бързото дискретно преобразуване на Фурие на всеки отрязък от сигнала:

$$S(k, l) = \sum_{n=0}^{N-1} s(n)W(n + lL) \exp\left(-\frac{j2\pi nk}{N}\right), 0 \leq n, k \leq N - 1 \quad (3.2)$$

където $S(k, l)$ е аудио сигналът, n е индексът на извадката във времевата област, N е броя на извадките на бързото дискретно преобразуване на Фурие, k е индексът на коефициентите на Фурие, l е относителното изместване на текущия аудио сегмент със стъпка L .

Получава се динамичният спектър за всяка 5-секундна част от аудио записа. За всеки сегмент се изчислява амплитудата на аудио спектрограмата, която в следващите стъпки на обработка се смята, че е изображение в сивата скала. Това позволява да се извърши допълнително потискане на шума във време-честотната област. Разглеждат се добре известни техники за обработка на изображения, които имат за цел да премахнат нецелевите звукови събития, като същевременно оставят непокътнати вокализациите на целевите видове. По-конкретно се използва 2-D итеративен процес на филтриране, базиран на морфологично филтриране на праговата аудио спектрограма.

Този метод има допълнителни стъпки за обработка на изображенията, приложени върху праговата спектрограма, за да се потисне остатъчният шум след прага чрез следното уравнение:



Фиг. 3.1: Цялостна блок-схема на предложения метод за автоматизирана параметризация на акустични събития на птичи песни

$$S_{\theta}(k, l) = \begin{cases} |S(k, l)|, & \text{ако } \theta_k \geq |S(k, l)| \\ 0 & \text{в останалите случаи} \end{cases}, \quad (3.3)$$

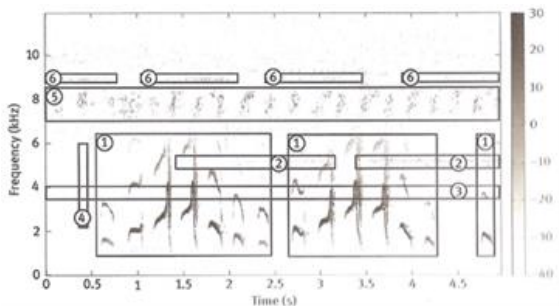
$$0 \leq k \leq N - 1, 0 \leq l \leq T - 1$$

Тук k и l са индексите на честотната лента и на аудио сегмента, N е броя на извадките на дискретното преобразуване на Фурие и T е общият брой аудио сегменти в спектрограмата. Прагът на спектрограмата θ_k се изчислява за всяка честотна лента k като средна стойност за цялата продължителност на спектрограмата:

$$\theta_k = a \frac{1}{T} \sum_{l=0}^{T-1} |S(k, l)|, \quad \begin{matrix} 0 \leq k \leq N - 1, \\ 0 \leq l \leq T - 1 \end{matrix} \quad (3.4)$$

Параметърът a позволява фина настройка на прага θ_k , така че прагът на спектрограмата да бъде оптимизиран по отношение на определена функция. Параметърът a може да бъде

зададен въз основа на предварителни знания или алтернативно може да се коригира автоматично. Процедурата за автоматизирана оптимизация зависи от наличието на набор от данни за разработка от няколко десетки спектрограми, които се използват при настройката на параметъра a . Трябва да се намери определен предварително дефиниран компромис между максимизиране на енергията на целевото събитие и минимизиране на енергията на нецелевите звукови събития .

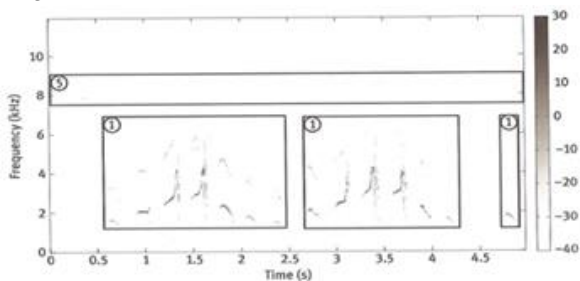


Фиг.3.2: Спектрограма на аудио сигнала след прилагане на праг на амплитудата[45]

$$S_{\theta_e}(k, l) = S_{\theta}(k, l) \odot C = \{z: C_z \subseteq S_{\theta}(k, l)\} \quad (3.5)$$

Формата и размерът на структуриращия елемент е специфичен за всяка различна задача. Резултатът от прилагането на уравнението (3.5) на праговата спектрограма (Фиг.3.2) е показана на Фиг.3.3. Както е показано на фигурата, след прилагане на оператора на ерозията се свиват ярките обекти, т.е. областите с висока енергия и се елиминират от спектрограмата маловажните петна, чиито размери са по-малки от размера на структуриращия елемент C .

Сравнявайки изображенията, показани на Фиг.3.2 и Фиг.3.3, се вижда, че в изчистената спектрограма на Фиг.3.3 във висока степен липсват конкурентните звукови събития, включително тези с етикет / 5 /, които са добре видими на изображението на Фиг.3.2.



Фиг.3.3: Праговата спектрограма след прилагане на оператора ерозия [45]

Прилага се оператора ерозия върху праговата спектрограма (3.3), с което се цели да получим спектрограма, почистена от детайли, по-малки от избрания структуриращ елемент C :

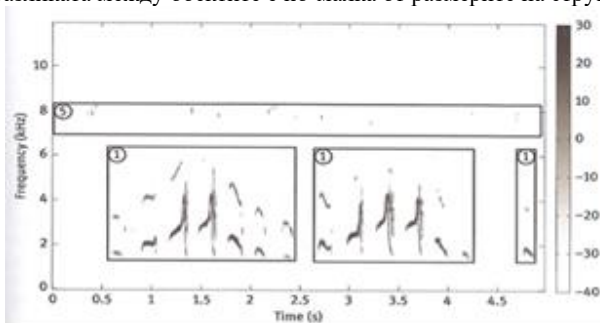
Звуковите събития, маркирани с /1/, са изчистени до известна степен, обаче те все още се различават ясно в изображението, което показва, че повечето от важната информация за тези вокализации са запазени непокътнати.

Установен подход за възстановяване на ерозиран контур е

операторът за дилатация, който запълва липсващите петна в структурата на обекта. Ако при прилагането на оператора за дилатация върху спектрограмата, ерозирана чрез уравнение (3.5), се използва същия структуриращ елемент \hat{C} , биха се възстановили обектите, оцелели от оператора на ерозията. Спектрограмата, която се получава след прилагане на оператора на дилатация (3.6):

$$S_{\theta_{ed}}(k, l) = S_{\theta_e}(k, l) \oplus \hat{C} = \{z: \widehat{C}_z \cap S_{\theta_e}(k, l) \neq \emptyset\} \quad (3.6)$$

ще има по-малък брой обекти от ерозираната спектрограма (Фиг.3.3). Тези обекти обикновено имат по-големи размери поради количественото увеличение в зависимост от формата и размера на структуриращия елемент \hat{C} . Следователно, прилагането на оператора за дилатация върху ерозираната спектрограма води до по-малък брой обекти поради сливането на обекти, разположени в непосредствена близост, тоест когато разликата между обектите е по-малка от размерите на структуриращия елемент \hat{C} .



Фиг.3.4:Праговата спектрограма след прилагане на оператора дилатация [45]

В общия случай формата и размерът на структуриращия елемент \hat{C} не са обвързани от избора на структуриращия елемент \hat{C} в уравнението (3.5). Дали двата

структуриращи елемента ще бъдат избрани еднакви или не е проблем, който

зависи от това дали трябва да се възстановят пикселите, ерозирани чрез (3.5) или също така се налага да се изглади или преструктурира обекта представляващ интерес в спектрограмата.

Резултатът от прилагането на уравнението (3.6) със структуриращ елемент $\hat{C}=\hat{C}$ върху ерозираната спектрограма (Фиг.3.3) е представен на Фиг.3.4. Както е показано на фигурата, остатъчният шум, останал след прилагане на оператора на ерозия (3.5), се усилва след прилагане на оператора за дилатация (3.6) - това е нежелан страничен ефект. Както е показано на Фигура 3.4, основната цел на оператора за дилатация е изпълнена успешно, като нулевите обекти, принадлежащи към звуковите събития, маркирани с /1/, са възстановени напълно и почти всички смущения от конкурентни звукови събития са премахнати. Изключение правят някои миниатюрни остатъчни петна в областта на етикета / 5 /. Тяхното съществуване е обусловено от фината настройка на прага (3.4) и от размера и остротата на структуриращите елементи в уравнения (3.5) и (3.6). Оптималният избор на тези свободни параметри на алгоритъма би гарантирал минимален остатъчен шум в морфологично филтрираната спектрограма.

В следващите стъпки на обработка се премахва остатъчният шум и се реконструират оригиналните обекти в морфологично филтрираната спектрограма (3.6). За тази цел се изпълняват следните стъпки на последваща обработка:

1. Оценява се прага θ_{ed} , пропорционален на средната стойност на всички пиксели в морфологично филтрираната спектрограма.
2. Прагът се нанася се върху морфологично филтрираната спектрограма, като се запазват елементите с енергия над прага и се нулират тези под прага.
3. Замянат се стойностите на запазените елементи с тези от оригиналната спектрограма $|S(k, l)|$.

$$\theta_{ed} = A \frac{1}{T} \sum_{l=0}^{T-1} \left(\sum_{k=0}^{N-1} S_{\theta_{ed}}(k, l) \right), 0 \leq k \leq N - 1, 0 \leq l \leq T - 1 \quad (3.7)$$

Прагът θ_{ed} се изчислява пропорционално на средната стойност на всички честотни елементи в морфологично филтрираната спектрограма, k и l съответстват на индексите на честотната лента и аудио сегмента, N е размерът на извадките на дискретното преобразуване на Фурие и T е общият брой на аудио сегментите в морфологично филтрираната спектрограма. Тук A е параметър, който може да бъде избран емпирично или коригиран автоматично въз основа на набора от данни за разработка. Всяка стойност на параметъра A в диапазона $[1, 20]$ обикновено осигурява добри резултати. Въпреки че стойността на A не е от решаващо значение за правилното функциониране на метода, тя също може да бъде коригирана автоматично с използването на определен алгоритъм за оптимизация.

След като бъде избрана стойността на прага θ_{ed} , може да се реализира праговата стойност и замяната на стойностите съгласно стъпки (2) и (3), както следва:

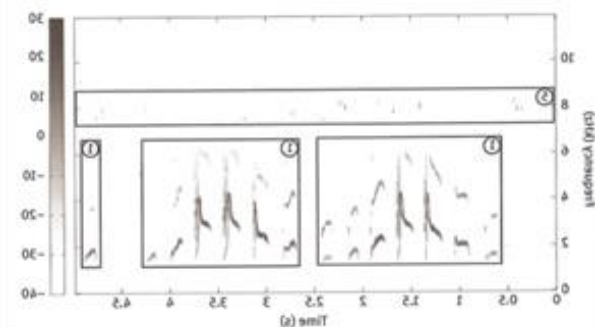
$$S_{\theta}(k, l) = \begin{cases} |S(k, l)|, & \text{ако } \theta_{ed} \geq S_{\theta_{ed}}(k, l) \\ 0 & \text{в останалите случаи} \end{cases}, \quad (3.8) \\ 0 \leq k \leq N - 1, 0 \leq l \leq T - 1$$

Резултатът от уравнение (3.8) е спектрограма (Фиг.3.5), която запазва целевите звукови събития, почти идентични с тези в оригиналната спектрограма, но въпреки това се възползва от морфологичното филтриране и стъпките за последваща обработка, които премахват конкурентните звукови събития.

Този метод за селективно морфологично филтриране използва някои предварителни знания за целевите звукови събития, които са включени главно чрез настройките на морфологичното филтриране, като например размера и формата на структуриращите елементи, но въпреки това този метод е приложим за широк кръг от излъчващи звук видове.

За някои методи за параметризиране на звука запазването на първоначалните свойства на звуковите събития (както тяхното местоположение, форма, така и енергия) е от решаващо значение, тъй като от това зависи представителността на изчислените аудио дескриптори. Това важи особено силно за изчисляването на параметрите на мел кеустрални коефициенти (MFCC) и за методи за параметризиране на аудио сигнали, които моделират времевото развитие на енергията в честотните ленти, представляващи интерес. Използването на метода за селективно морфологично филтриране, води до спектрограма, която има много нули в колоните и редовете и следователно MFCC няма да доведе до ефективно представяне на акустичните събития. Въпреки това, някои 2-D методи за параметризиране на звука не зависят от запазването на цялата информация за

оригиналните звукови събития и значителна част от информацията може да бъде елиминирана чрез намаляване на общата сложност. Например, това се отнася за случая, когато етапът на извличане на аудио функция се възползва само от 2-D контурите на звуковите събития.



Фиг.3.5. Спектрограма след прилагане на праг върху морфологично филтрирана спектрограма [45]

Методът за морфологично филтриране не е настроен за изчисляване на определен набор от аудио дескриптори. Вместо това може да служи като предварителна обработка на аудио сигнала за различни алгоритми за параметризиране на звука, сред които кепстрални коефициенти,

дескриптори базирани на уейвлет пакети или други 1-D аудио дескриптори. Освен това методът е съвместим с оценката на базирани на изображения дескриптори въз основа на области, които представляват интерес в спектрограмата и имат за цел да открият двузмерни форми и контури в изчистената спектрограма.

Прагът на спектрограмата се осъществява чрез прилагане на специфични за обхвата прагове на всяка колона на спектрограмата. Използването на морфологично филтриране с операторите за дилатация и ерозия води до спектрограма, която съдържа по-малък брой обекти. Това се дължи на сливането на обекти, разположени в непосредствена близост, т.е. когато разликата между обектите е по-малка от размерите на структуриращия елемент и елиминирането на малки обекти, които са разположени далеч от други обекти. Като крайна стъпка на предварителна обработка се извършва допълнително ограничаване на прага, за да се премахнат пикселите с най-ниска енергия в изображението в сивата скала. Прагът е избран като 10 пъти средната стойност, изчислена за всички пиксели в изображението.

3.4 Параметризиране чрез интерполация със смес от Гаусови модели

Предложеният метод за параметризиране чрез интерполация със смес от Гаусови модели включва няколко етапа на обработка, които прилагат сегментиране на птичата вокализация като акустични събития, получаване на характеристичната крива за всяко акустично събитие и интерполация на кривата с модел, базиран на GMM. По-конкретно, стъпката на сегментиране използва изображението на сивата скала. Изборът на области в спектрограмата, които съдържат видни акустични събития, се получава след праг на морфологично изчистеното изображение на сивата скала с адаптивен праг. Този праг се изчислява статистически въз основа на хистограмата на средните стойности в изчистеното изображение на сивата скала, получено за тази цел. Тук задаваме прага на 30% от максималната стойност в хистограмата.

Етапите на обработка, включени в оценката на характеристикната крива за акустично събитие, се състоят от:

- Откриване на честотния компонент с най-висока енергия във всяка колона на обработената спектрограма.
- Изглаждане на доминиращия контур на честотата.
- Откриване на границите между сричките и сегментиране на последователности от 1-D стойности.

По време на оценката на характеристикната крива правим преход между 2-D представяне на информацията във времево-честотното пространство и 1-D представяне на доминиращия честотен контур, характеризиращ се с времеви ред от стойности. По-конкретно, по време на стъпка на обработка 1, т.е. откриване на честотния компонент с най-висока енергия във всяка колона на спектрограмата, методът работи директно върху елементите на обработеното изображение в сива скала, които се определят от триплета (амплитуда, честота, време). Резултатът е матрица, която има само един ненулев елемент се прилага за коригиране на удвояването и намаляването наполовина на оценката на доминиращата честота поради хармониците с висока амплитуда. След това се използва филтър за изглаждане на характеристикната крива. В резултат на това се получава плавна крива, която се характеризира с времеви ред от стойности, дефинирани от дуплет (честота, време).

Изгладената характеристична крива претърпява обработка с правоъгълна прозрачна функция от 50 ms. Те се елиминират чрез малки прекъсвания в пиковата енергийна траектория поради грешки при нейното определяне или от наличието на съпътстващи акустични събития от други източници. Области, в които липсва ясно изразена вокализация са преобразувани в дълги последователности от нули. В резултат на прилагания алгоритъм две акустични събития са разделени като отделни, ако са на разстояние повече от 100 ms. Части от характеристикната крива, в които няма събитие повече от 100 ms, се считат за различни събития и се интерполират индивидуално.

На този етап трябва да се изясни, че контурите могат да имат дължина до стотици стойности и, че тяхната дължина се различава значително между типовете срички. Освен това дължината на контура за един и същ тип сричка варира при различните реализации от един и същи тип песен. Това мотивира да се търси единно и компактно представяне на информацията, кодирана във всеки контур чрез някакво параметрично моделиране на контура. Разглежда се GMM-базирана интерполация на краткосрочния енергиен контур на доминиращата честотна компонента в акустичния сигнал. Такъв модел осигурява необходимото компактно представяне на информацията и демонстрира благоприятни характеристики в сравнение с други методи за моделиране на криви. Анализът на резултатите, получени за различните видове характеристикни криви, показва, че методът за интерполация, базиран на GMM с M на брой компоненти, осигурява подходящо представяне на ключовите характеристики без значителна загуба на информация:

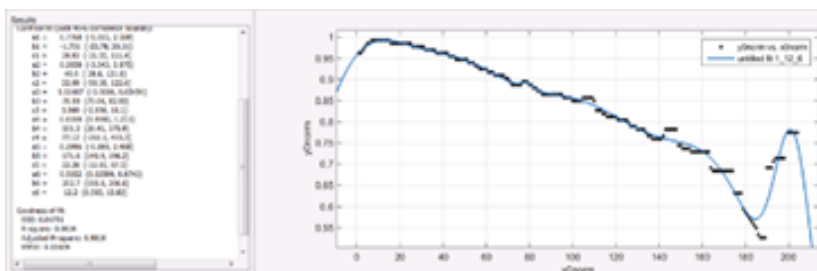
$$f(x) = \sum_{n=1}^M a_n e^{-\left(\frac{x-b_n}{c_n}\right)^2} \quad (3.9)$$

където центровете b_n , разпръскванията c_n и амплитудите a_n са параметрите на М Гаусови компоненти. Такъв модел има общо $3 \cdot M$ регулируеми параметри. В сравнение със стотиците стойности в оригиналния контур, тези $3 \cdot M$ стойности съответстват на многократно намаляване на количеството данни.

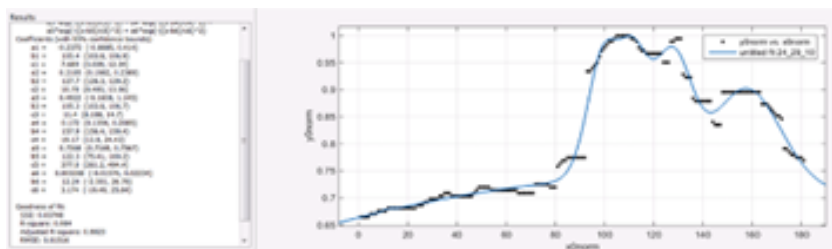
3.5 Експериментално изследване за случая на *Mymotherula Multostriata*

Представен е пример за аудио запис на *Mymotherula multostriata*, направен в естественото местообитание на вида. Предварително обработения сигнал е сегментиран, за да се отделят изолирани акустични емисии, следвайки процеса, обсъден по горе, който доведе до общо 78 сегмента, всеки описан с характеристична крива. Сегментите, които имат сходни форми са групирани в девет широки категории.

На Фиг.3.6 и Фиг.3.7 са представени примери за интерполация на характеристичната крива с модела GMM.



Фиг.3.6: Интерполация на базата на GMM



Фиг. 3.7: Интерполация на характеристична крива с GMM модел

ИЗВОДИ към Глава III

Въз основа на изложението и експерименталната валидация могат да се направят следните изводи:

Синтезиран е метод за автоматизирано извличане и параметризиране на характеристичната крива на акустични събития на птичи вокализации, които позволява акустичните събития с различна дължина и сложност да бъдат представени с фиксиран брой параметри.

Получена е експерименталната валидация на предложения метод с реални записи на акустичните емисии на птичия вид *Mugmotherula multostriata*, регистрирани в полеви условия. Показано е, че параметрите на предложения метод за GMM интерполация се потвърждават като подходящи за описване на характеристичната крива.

Анализът на параметрите, получени за 78 аудио сегмента, позволява да се направи заключение, че тези параметри имат свойства, които ги правят подходящи да служат като елементи на характеристичен вектор в приложения за разпознаване на акустични емисии на биологичните видове.

ГЛАВА IV. МОДЕЛИРАНЕ НА ПТИЧИТЕ ВОКАЛИЗАЦИИ

В настоящата глава от дисертационния труд представяме метод за автоматизирано разпознаване на типа срички на птичите вокализации. За тази цел автоматично се сегментира песента на птиците на акустични събития и впоследствие всеки сегмент се моделира чрез интерполация, базирана на GMM, на енергията на доминиращия честотен компонент. След това параметрите на модела на GMM се подават към класификатор, за да се разпознае типа на сричките на птичата вокализация. Оценена е практическата стойност на този подход, като се използват публично достъпни полеви записи на видове *Mugmotherula multostriata*, които са били регистрирани в естествени местообитания. Експерименталният протокол се основава на седемдесет и осем акустични събития, които са получени след автоматичното сегментиране на аудио сигнала. Отчита се точност на разпознаване до 98%, в зависимост от метода на класификация и конкретния тип сричка. Обобщавайки експерименталните резултати, получени в това изследване, се стига до заключението, че предложеният метод има добър потенциал за постигане на по-висока точност на разпознаване, но е необходима допълнителна работа за използването му от практически приложения.

4.1. Въведение в проблема

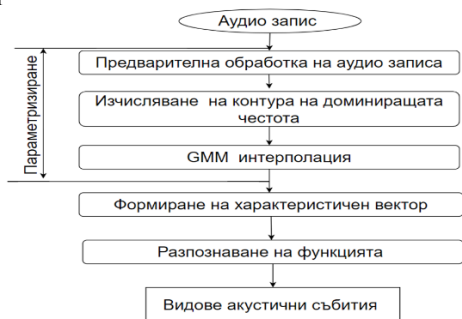
Характеристичната мелодия на птичи песни обикновено се получава въз основа на енергията на доминиращата честотна компонента в акустичния сигнал. Характеристичната мелодия на вокализацията на птици е сред най-информативните дескриптори за акустичното моделиране и автоматичното разпознаване на видове птици въз основа на техните акустични емисии.

Представят се нови методи за автоматизирано моделиране на характеристичната мелодия на птичи песни чрез интерполация на време-честотния контур на енергията на доминиращия честотен компонент. Доминиращият енергиен контур обикновено се получава чрез автоматичен метод за сегментиране на птичите песнопения. Резултатите от такова сегментиране са акустични събития, които тук грубо се наричат срички на птичите песни. По-конкретно, се счита, че всяко отделно акустично събитие може да бъде представено чрез неговия доминиращ енергиен контур - характеристичната му крива. Понякога тези криви на специфични срички на птичи песни са много сходни, дори еднакви сред различните видове, но динамичните поредици от тях, т.е. характеристичната мелодия е уникална за всеки конкретен вид.

4.2. Метод

На Фиг.4.1 е представена цялостната концепция на предложения метод за автоматизирано разпознаване на вида на акустичните събития. Предполага се, че аудио запис с продължителност най-малко няколко секунди, който съдържа вокализация на

целевия вид птици се подава на входа към предварителната обработка на звука. В полевите записи на птици, аудио сигналът често съдържа съпътстващи звуци от други видове, като бозайници, насекоми или земноводни, които спонтанно и независимо един от друг вокализират. За да се запази за по-нататъшна обработка само честотния диапазон, който представлява интерес за даденото изследване, по време на стъпката за предварителна обработка на звука, първо се прилага лентов филтър. Тук се разглежда лентов филтър Butterworth от четвърти ред с честотна лента 2-10 kHz. Това до голяма степен потиска енергията, дължаща се на вятъра в микрофона, както и на звуците на бозайници и някои земноводни. След това сигналът се сегментира с 16-ms прозорец на Хаминг, за да се образуват припокриващи се сегменти и след това всяка порция сегменти се трансформира в честотната област с помощта на дискретна трансформация на Фурие (DFT). За целите на по-нататъшната обработка се подреждат Фурие коефициентите, изчислени за тези сегменти, докато се образува 10-секундна част от сигнала. След това се изчислява спектрограмата, която по-долу се счита за изображение в сива скала. За да се елиминират конкурентни акустични събития от насекоми и други животински видове, които се припокриват с честотния диапазон на целевия птичи вид, се прилага итеративен алгоритъм за потискане на шума, базиран на морфологично филтриране на спектрограмата. В резултат на това се намалява риска спектрограмата да съдържа къси сегменти с пикове на енергия или дълги стационарни звукови събития от насекоми или земноводни, които могат да попречат на сегментирането на птичата вокализация на срички.



Фиг.4.1: Цялостна концепция на предложения метод за автоматизирано разпознаване на категориите акустични събития

След като основните източници на смущения бъдат потиснати, се продължава със стъпката за параметризиране на звука, която има за цел да разложи записа на срички и да приложи

параметризацията, базирана на GMM. За целта първо се търси пиковата енергия във всяка колона на спектрограмата. В резултат на това се получава суров контур от енергийни пикове, който да проследи по-голям обсег, свързан с доминиращата честотна компонента. Участъците с енергиен пик под определен предварително зададен праг са настроени на нулева енергия. Генерира се ново черно-бяло изображение, където черните точки отговарят само на позициите, където енергията е над гореспоменатия праг. За да се елиминират отклоненията в единична и двойна точка от траекторията на контура, се прилага процедура за изглаждане, която има целенасочено дефинирани правила за корекция на грешки, последвани от конволюция. По-конкретно, за да се получи по-

гладък контур, изчисляваме сумата на конволюцията между контура и правоъгълна прозрачна функция с продължителност W . Стойността на W е специфична за видовете и обикновено се основава на познанията в областта. В противен случай трябва да се определи статистически въз основа на някои записи. В нашия случай за видовете *Mugmotherula multostriata* приемаме $W = 50$ ms.

В крайна сметка контурът се сегментира на срички или по-кратки аудио събития в зависимост от дължината на интервалите между различните му части. Когато две съседни части на контура се намират между някакво предварително определено разстояние θ , ние ги разглеждаме като последователност и ги съхраняваме заедно в един сегмент. Когато разстоянието е по-голямо от прага θ , разделяме тези части като отделни срички. В конкретния случай считаме $\theta = 100$ ms. Като цяло стойността на θ е специфичен за даден вид параметър и трябва да се коригира въз основа на статистически анализ или въз основа на специфични за дадена област знания.

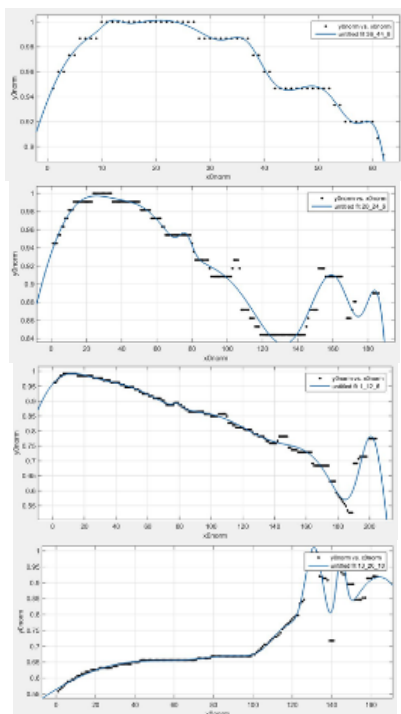
Контурите могат да имат дължина до стотици рамки и дължината им се различава значително между типовете срички. Освен това дължината на контура за един и същ тип сричка варира при различните реализации от един и същи тип песен. За тази цел се търси компактно представяне на информацията, кодирана във всеки контур чрез някакво параметрично моделиране. Тук се разглежда GMM-базирана интерполация на енергийния контур на доминиращата честотна компонента в акустичния сигнал. GMM осигуряват адекватно моделиране на контурите на сричките, когато се състои от шест Гаусови компонента. Такъв модел осигурява необходимото компактно представяне на информацията и демонстрира благоприятни характеристики в сравнение с други методи за моделиране на криви. Анализът на резултатите, получени за различните видове характеристични криви показва, че методът за интерполация, базиран на GMM, с шест компонента осигурява подходящо представяне на ключовите характеристики без значителна загуба на информация. Такъв модел има 18 регулируеми параметри – центровете b_n , разпръскванията c_n и амплитудите a_n на шестте гаусови компонента. В сравнение със стотиците стойности в оригиналния контур, тези 18 стойности съответстват на многократно намаляване на количеството данни.

$$f(x) = \sum_{n=1}^6 a_n e^{-\frac{(x-b_n)^2}{c_n}} \quad (4.1)$$

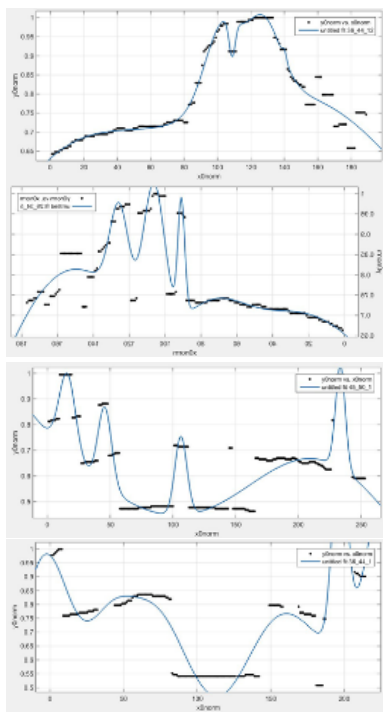
На Фиг.4.2 и Фиг.4.3 показваме различни примери за характеристични контури, получени за различни типове срички, налични в избраните записи на *Mugmotherula multostriata*, заедно с контурите на техните GMM-базирани интерполации. Контурите, показани на Фиг.4.2, могат да бъдат характеризирани с бавна динамика и с по-ясно съвпадение между действителните точки и кривата, дефинирана чрез интерполация, базирана на GMM.

Контурите, показани на Фиг.4.3, се характеризират с по-висока степен на променливост и бързи динамични промени в амплитудата, което отразява по-неточно действителните точки на контура. Предполага се, че това не е съществен проблем, тъй като цялостната динамика и тенденцията на контура се улавят адекватно и следователно разпознаването на типа сричка няма да бъде засегнато значително. След като

параметрите за интерполация на GMM се коригират, така че да получим адекватно напасване на точките в контура, параметрите на модела се подреждат заедно, за да се образува характеристичен вектор , който се подава към етапа на класификация.



Фиг.4.2: Различни характеристични контури, получени за различни типове срички, налични в записите на *Myrmotherula multostriata*



Фиг. 4.3: Примери за характеристични контури с висока вариабилност, получени за различни типове срички, налични в записите на *Myrmotherula multostriata*

Изпълнението на различните видове задачи за разпознаване на звукови събития, се основава на 1-D или 2-D аудио параметризация и някои методи за класификация. Откриването на определен тип сричка за нуждите на експеримента в настоящата дисертация може да се осъществи с помощта на всеки дискриминационен, генеративен или хибриден метод на класификация.

4.3.Настройките на експеримента

Методът, представен по-горе, е оценен в експериментален протокол, основан на публично достъпен аудио запис на видовете птици *Myrmotherula multostriata*, записани в местообитанието им в дивата природа. Последвани са стъпките за сегментиране, описани в предходната глава , за да се обработи едноминутен аудио запис. В резултат на това се получават седемдесет и осем сегмента с акустични събития на целевите видове, които в

по-широк смисъл наричаме срички. Всички сегменти са обработени, за да се получи характеристичния контур и интерполацията, базирана на GMM. Оптималните стойности на 18-те GMM параметъра, получени за определен сегмент, се считат за характеристичен вектор. По този начин, в резултат на сегментирането и параметризирането на наличния запис, се получава матрица с размерност 78 x 18, която съдържа 78 характеристични вектори, всеки с размерност 18. Тези данни са разделени в девет широки групи в зависимост от типа на сричката. Груповите етикети могат да се използват като истински етикети, когато се изисква автоматизирано идентифициране на типа сричка.

Проведени са валидационни експерименти за три от тези групи (обозначени като Група 1, Група 2 и Група 3), които съдържат достатъчен брой примери за целевия тип сричка. Последното се изисква за надеждното обучение на статистическите модели на специфичните за срички детектори. Във всички експерименти тези детектори разпознават една от целевите групи спрямо всички останали типове срички, така че намаляваме проблема до двоично - открита целева сричка: „true“ или „false“.

Всеки детектор е реализиран с две различни техники за класификация: (1) с многослойна перцептронна невронна мрежа (MLP NN) и (2) с мрежа на Бейс (Bayes Network). За да се настроят техните регулируеми параметри, се използва подхода за търсене на мрежа, целящ постигане на възможно най-висока точност на разпознаване. Във всички случаи автоматизираното разпознаване се основава на метода за разделяне на данните.

$$Accuracy = \frac{1}{2} \left(\frac{n_{cf}}{n_f} + \frac{n_{ct}}{n_t} \right) \cdot 100 \% \quad (4.2)$$

където n_{cf} (истински отрицателен) е броят на коректно откритите екземпляри, свързани с категория „false“, и n_{ct} е броят на коректно откритите вектори за категория „true“. Тук n_f и n_t са общият брой случаи, свързани с отрицателните и положителните категории във всеки експеримент.

4.4. Експериментални резултати

Експерименталните резултати за трите типа срички на два класификатора са представени в Таблица 4.1. Всяка клетка в таблицата показва точността на класификацията в проценти за специфичния за сричките детектор. Както е показано в Таблица 4.1, нито един от класификаторите не предоставя надеждни резултати от разпознаването - за определен тип сричка единият или другият класификатор не успява да изгради стабилни статистически модели, което води до неприемливо ниска точност на класификация за тази група. Детекторът, базиран на мрежата на Бейс, представя неочаквано ниска точност на класификация при разпознаване на сричките от група 1, а детекторът, базиран на MLP NN, показва ниска точност на класификация за сричките от група 2 и група 3. Разширявайки анализа, като се пренебрегва най-ниската точност на класификация за всеки от детекторите, можем да кажем, че резултатите от класификацията са обещаващи. Счита се, че има добра перспектива за подобряване на точността на класификацията, когато са налични повече записи за обучение на специфичните за сричките детектори.

Таблица 4.1: Класификационна точност на специфични сричкоразпознаващи детектори изпълнени с MLP NN и Bayes Network.

	Група 1 Точност[%]	Група 2 Точност [%]	Група 3 Точност [%]
MLP NN детектор	83.0 %	56.4 %	75.8 %
Bayes Network детектор	66.7 %	90.3 %	98.4 %

ИЗВОДИ към Глава IV

В настоящото изследване се оценява хипотезата, че характеристичната крива на сричките на птичи песни може да се използва успешно като описателна характеристика в автоматизирани методи за откриване на типа сричка. Освен това е предложен метод, основан на GMM, за компактно представяне на характеристичния контур на различни типове срички, които имат различна продължителност и динамика. Експерименталната оценка на предложения метод със записи на видовете птици *Murmotherula multostriata* показва, че параметрите на интерполацията, базирана на GMM се потвърждават като подходящи характеристики за разпознаване на типовете срички. Точността на класификацията, отчетена в тази работа, не е достатъчна за практически приложения, но се счита, че предложеният метод има потенциал за подобряване на производителността, когато са налични повече данни. Бъдещите изследователски усилия ще имат за цел да изследват приложимостта на предложения метод и за други видове пеещи птици.

ЗАКЛЮЧЕНИЕ

От получените експериментални резултати и направените изводи се стига до следните заключения:

Необходими са продължаващи изследвания на методите и техническите средства за анализ и разпознаване на акустичните емисии на биологичните видове. Изисква се използването на съвременни теоретични достижения, модерни апаратни средства, хардуер, софтуер и средства за машинно обучение

В дисертационния труд е разработен технологичен подход, при който поетапно се диференцират, анализират и изследват характеристичните мелодии на птичите вокализации, базирани на множество комбинации от методи и признаци, основаващи се на обективни звукови емисии.

Предварителната обработка включва изчисления със специфичен характер върху предоставените цифрови данни, с цел получаването на начални данни за първично редуциране на описателите, диференциране и сегментация на сигнала. Извършва се време-честотно разлагане и извличане на характеристиките на сигналите. Процесът завършва с последваща обработка, разпознаване и класификация.

Като се има предвид важността на получените данни и тяхното практически приложение е направен сравнителен анализ по отношение на критерия точност, което допълнително потвърждава достоверността на получените резултати.

ПРИНОСИ ПО ДИСЕРТАЦИЯТА

Научно-приложни приноси

1. Създаден и валидиран е метод за определяне траекторията на доминиращия честотен компонент, определящ характеристикната мелодия на птичи вокализации с ясно изразена хармонична съставна, регистрирани в дивата природа.
2. Създаден и валидиран е метод за параметризиране на характеристикната мелодия на птичи вокализации, използващ смеси от Гаусови функции (Gaussian Mixture Models, GMM), за да интерполира отделните сегменти от характеристикната крива, изчислена за всяко акустично събитие.
3. Създаден и валидиран е метод за автоматизирано разпознаване на типовете срички на птичите вокализации, при който се постига точност до 98%.

Приложни приноси

4. Създадено е програмно осигуряване, реализиращо предложените методи.
5. Проведени са експериментални изследвания за определяне на оптималните параметри на моделите.
6. Проведени са числови симулации за оценка работоспособността на предложените методи.

ПУБЛИКАЦИИ ПО ТЕМАТА НА ДИСЕРТАЦИОННИЯ ТРУД

1. N.Dukov, **P.Deneva**, T.Ganchev, P.Parushev, “Intelligent audio logger for bird sound“, Списание „Акустика“ 2016, ISSN 1312-4897, стр.87-90.
2. **Денева, П.С.**, “Разпознаване на характеристичната мелодия на биоакустични сигнали“, Списание „Акустика“ 2017, ISSN 1312-4897, стр.57-60.
3. **Денева, П.С.**, Ганчев,Т., “Оценка на грешката при моделиране на характеристична мелодия на птичи вокализации“, Списание „Акустика“ 2018, ISSN 1312-4897, стр.12-17.
4. **Денева, П.С.**, Ганчев,Т., “Моделиране на характеристична мелодия на птичи вокализации“, ТУ-Варна, КНТ 2018, ISSN 1312-3335, стр.156-160.
5. **P. Deneva**, T.Ganchev, “GMM-based Parametrization of the Characteristic Melody of Bird Vocalizations“, Twenty First International Conference of Geometry, Integrability and Quantization, 2019, ISSN 1314-3247,pp 118-126.
6. **P. Deneva**, T. Ganchev, „Automatic recognition of the characteristic melody of bird vocalizations“, International Conference on Biomedical Innovations and Applications, 2019, <http://biaconf.tu-varna/index.php,10.1109/BIA48344.201>.
7. **П. С. Денева**, “Система за цифров анализ на звукови записи на птичи вокализации“, ТУ-Варна, КНТ 2021, ISSN 1312-3335, стр.71-77.

Благодарности на:

Желая да благодаря специално на моят научен ръководител, професор Тодор Ганчев, за вдъхновението, което ми предаде, за търпението му, както и за съветите и насоките, които ми предоставяше.

Благодаря на семейството ми, което неизменно бе до мен през тези години и ми вярваше, че ще постигна успех.

Бих желала да благодаря и на всеки, който се е запознал и прочел този труд. Благодаря за съветите, оценките и мненията по проблемите в тази дисертация.

ABSTRACT

**Dissertation Title: Research On Automatic Recognition Of Biological Species From
Their Acoustic Emissions
of the Requirement for the Degree Doctor of Philosophy
by Prolet Stoycheva Deneva**

The main goal of the dissertation is to develop methods for modelling and automatic recognition of acoustic emissions from birds, mammals, reptiles and insects in their natural habitats.

Chapter I presents the specifics and challenges of the implementation of automatic recognition of vocalizations of different species. An overview of the available solutions to the problem and how the results are applied is made. The possibilities for implementing the set task with the different technologies for realization are analyzed and researched.

Chapter II of the dissertation explores the possibility of determining the characteristic melody by interpolation through the spectral peaks of the fundamental harmonic of the signal, determined for short time intervals. The conceptual solution to the problem is presented, and a method for finding the trajectory of the characteristic melody in bioacoustic recordings of birds registered in the wild is proposed. Numerical results from the application of the method are shown, and a summary, conclusions, and specifications of the potential applicability are synthesized.

Chapter III presents a new method for parameterizing the characteristic melody of bird vocalizations, which uses Gaussian Mixture Models (GMM) to interpolate the individual segments of the characteristic curve calculated for each acoustic event. This process results in a parametric model that describes segments of different lengths and complexity with a predetermined number of parameters. Experimental studies show that modelling with a six-component GMM provides the necessary precision and accuracy to represent the characteristic melody. The parameters obtained through this model can serve well in modelling bird vocalizations or species recognition applications.

Chapter IV of the dissertation presents a method for automated recognition of syllable types of bird vocalizations. For this purpose, the song of the birds is automatically segmented at acoustic events, and subsequently, each segment is modelled by GMM-based interpolation of the energy of the dominant frequency component. The GMM model parameters are then submitted to a classifier to identify the syllable type of the bird's vocalization. The practical value of this approach was assessed using publicly available field records of *Myrmotherula multostriata* species that have been recorded in natural habitats. The experimental protocol is based on seventy-eight acoustic events obtained after automatic audio signal segmentation. The recognition accuracy of up to 98% is reported, depending on the classification method and the specific syllable type. Summarizing the experimental results obtained in this study, it is concluded that the proposed method has good potential for achieving higher recognition accuracy, but further work is needed to use it in practical applications.

