

КОМПЬЮТЕРНЫЕ НАУКИ И ИНФОРМАТИКА

Научная статья

УДК 04.934.2

DOI: 10.17072/1993-0550-2024-4-104-116

<https://elibrary.ru/ENFGRE>**Применение математических методов
в искусственной нейронной сети для распознавания
в голосе "фальшивых" эмоций****Иван Владимирович Быков¹, Наталья Юрьевна Ротанева²,
Александр Петрович Шкарапута³**^{1,3} Пермский государственный национальный исследовательский университет, г. Пермь, Россия² Мариупольский государственный университет имени А.И. Куинджи, г. Мариуполь, Россия¹vanka.perm@mail.ru²n.rotaneva@mgumariupol.ru³shkaraputa@psu.ru

Аннотация. В данной статье предложена методика сравнения эмоциональной составляющей голосов спикеров, реально испытывающих эмоции, и тех, которые пытаются их воспроизвести. Методика предполагает использование в качестве классификаторов отношение частот локальных максимумов спектра звуковой волны. В данной работе исследовались 4 категории базовых эмоций: гнев, грусть, страх, радость. Для получения сравнительных характеристик в каждой категории исследовалось по 30 записей с "фальшивой" и "истинной" эмоцией – всего 240 записей. Был проведен статистический сравнительный анализ классификаторов и обнаружены существенные различия в данных для искренних эмоций и их имитации. Также, была спроектирована модель искусственной нейронной сети, на основе которой создана программа для распознавания эмоционального посыла, содержащегося в записях голоса человека. На обученной искусственной нейронной сети проведен эксперимент по определению эмоционального состояния спикера и истинности или фальшивости его эмоционального посыла. Тестирование звуковых файлов с различными эмоциями показало хорошие результаты как для распознавания самого эмоционального состояния говорящего, так и истинности его эмоционального посыла.

Ключевые слова: распознавание эмоций; фальшивые эмоции; базовые эмоции; музыкальный интервал; классификаторы эмоционального состояния; просодия

Для цитирования: Быков И.В., Ротанева Н.Ю., Шкарапута А.П. Применение математических методов в искусственной нейронной сети для распознавания в голосе "фальшивых" эмоций // Вестник Пермского университета. Математика. Механика. Информатика. 2024. Вып. 4(67). С. 104–116. DOI: 10.17072/1993-0550-2024-4-104-116. <https://elibrary.ru/ENFGRE>.

Статья поступила в редакцию 25.10.2024; одобрена после рецензирования 15.11.2024; принята к публикации 01.12.2024.



Эта работа © 2024 Быков И.В., Ротанева Н.Ю., Шкарапута А.П. распространяется под лицензией CC BY 4.0. Чтобы посмотреть копию этой лицензии, посетите <https://creativecommons.org/licenses/by/4.0/>

COMPUTER SCIENCE

Research article

Mathematical Methods Application in an Artificial Neural Network to Recognize "Fake" Emotions in the Voice

Ivan. V. Bykov¹, Natalya. N. Rotaneva², Alexander. P. Shkaraputa³

^{1,3} Perm State University, Perm, Russia

² Mariupol State University named after A.I. Kuindzhi, Mariupol, Russia

¹vanka.perm@mail.ru

²n.rotaneva@mgumariupol.ru

³shkaraputa@psu.ru

Abstract. This paper proposes a methodology for comparing the emotional component of the voices of speakers actually experiencing emotions and those trying to reproduce them. The technique assumes the use of as classifiers the ratio of frequencies of local maxima of the spectrum of the of a sound wave. In this work we studied 4 categories of basic emotions: anger, sadness, fear, joy. To obtain comparative characteristics in each category 30 records with "fake" and "true" emotion were investigated – 240 records in total. A statistical comparative analysis of the classifiers was performed and found a significant differences in the data for the genuine emotion and its imitation. Also, a model of an artificial neural network, based on which a program was created to recognize the emotional message contained in human voice recordings. On the trained artificial neural network, an experiment was conducted to determine the emotional state of the speaker and the truth or falsity of his emotional message. Testing of sound files with different emotions showed good results for recognizing both the emotional state of the speaker and the truthfulness or falsity of the speaker's emotional message.

Keywords: *emotion recognition; fake emotions; basic emotions; musical interval; emotional state classifiers; prosody*

For citation: Bykov, I.V., Rotaneva, N.Yu. and Shkaraputa, A.P. (2024), "Mathematical Methods Application in an Artificial Neural Network to Recognize "Fake" Emotions in the Voice", *Bulletin of Perm University. Mathematics. Mechanics. Computer Science*, no. 4(67), 104-116. DOI: 10.17072/1993-0550-2024-4-104-116. <https://elibrary.ru/ENFGRE>.

The article was submitted 25.10.2024; approved after reviewing 15.11.2024; accepted for publication 01.12.2024

Введение

Системы распознавания человеческих эмоций – одни из самых востребованных и быстро развивающихся программно-аппаратных комплексов. Они востребованы во многих сферах человеческой деятельности, от игровой индустрии до выявления криминала. При этом определение искренности эмоции может играть большую роль в работе спецслужб, при прослушивании телефонных разговоров, а также допросах, при оформлении кредита или собеседования при приеме на работу.

В настоящее время для выявления лжи используются различные инструменты. Наиболее известным инструментом является полиграф – устройство, позволяющее осуществлять инструментальные психофизиологические исследования на основе записи показателей дыхания, электрического сопротивления кожи, кровообращения и других физиологических параметров.

Голосовой детектор лжи "RiskControl" использует другой принцип работы.

Это уже алгоритм, определяющий требуемое человеку время для ответа на вопрос.

Он также оценивает продолжительность ответа, громкость и спектр голоса, сопоставляя их с эталонными значениями. Например, при длительной паузе повышается вероятность того, что собеседник говорит неправду, так как ему требуется время для придумывания ответа. Громкий разговор может говорить о нервозности человека. [1]

Существует детектор лжи, который основан на определении миографического тремора мышц. Он анализирует звукозаписи речевых высказываний и выявляет не воспринимаемые на слух акустические особенности голоса, вызванные стрессом. [2]

В данной работе предлагается определять искренность человека по его эмоциональному послы, что позволяет не только определять ложность его высказываний, но также определять наигранность эмоций. Это, в свою очередь, может расширить круг решаемых задач.

Американский психолог Полу Экман выделил семь базовых эмоций [3]: радость; грусть; гнев; отвращение; презрение; страх; удивление. В данной работе будут исследоваться четыре базовые эмоции: гнев, грусть, страх, радость. Выбор этих эмоций обусловлен тем, что для них легче всего найти соответствующие звуковые записи и их проще разделить на экспертном уровне. Записи, содержащие "истинные" и "фальшивые" эмоции брались из разных источников: видеороликов, фильмов с актерами, записями аудиокниг... Иногда в записях не было даже слов, а например был только смех или крик.

В работе [4] показано, что метод отношения частот локальных максимумов спектра звуковой волны может быть применен для распознавания эмоционального состояния человека, а сами отношения хорошо соотносятся с музыкальными интервалами, такими как терция, квинта... В отличие от других подходов [5], например, метода мел-частотных кепстральных коэффициентов (MFCC) [6], данный метод не зависит от многих таких факторов, как паузы, высота тона, и, даже, языка говорящего [7, 8]. Суть метода заключается в том, что в спектре звуковой волны выделяют несколько локальных максимумов амплитуд. Затем частоты, которые им соответствуют, делятся попарно одна на другую, таким образом получают классификаторы. Например, в записях с радостной эмоцией, часто встречается отношение 0,84, а оно, в свою очередь соответствует музыкальному интервалу – большая терция. Большая терция в теории музыки – это светлый, веселый, мягкий консонанс. В записях с грустной эмоцией часто встречается отношение 0,79 – мягкий, нежный, грустный консонанс.

На рис. 1 показан спектр звуковой волны, полученный на основе дискретного преобразования Фурье, с выделенными локальными максимумами.

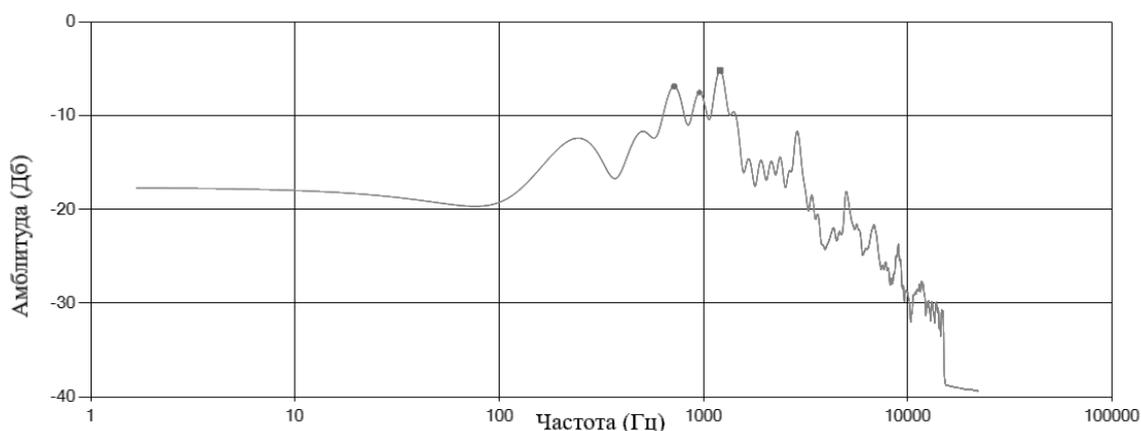


Рис. 1. Пример нахождения 3-х локальных максимумов амплитуд

Ниже в таблице 1 приведены отношения частот, соответствующие этим амплитудам.

Таблица 1. Полученные отношения частот

Частота (Гц)	1203	718	955
1203	1	1,69	1,27
718	0,59	1	0,75
955	0,79	1,33	1

Здесь полужирным выделены отношения, соответствующие музыкальным интервалам: малая терция и чистая квинта.

Постановка задачи

Задача заключается в том, чтобы на основании построенных классификаторов научиться различать как эмоциональное состояние говорящего человека, так и искренность его эмоционального посыла. Для решения данной задачи потребовалось пройти несколько этапов, на каждом из которых требовалось решить свою подзадачу. Так, на начальном этапе нужно собрать необходимую библиотеку звуковых файлов с соответствующими эмоциями. Сами записи на экспертном уровне (решением не менее трех человек) необходимо определить в соответствующую категорию по типу эмоции и искренности эмоционального посыла. В дальнейшем необходимо создать программу, на основе которой, с помощью дискретного преобразования Фурье, можно построить спектр звуковой волны и определить локальные максимумы амплитуд. После этого этапа необходимо получить набор классификаторов (отношений частот) для каждой записи. Когда классификаторы получены, то можно переходить к их сравнению на основе статистического анализа и нахождению отличий. Так как методы искусственных нейронных сетей являются, в свою очередь, частным случаем статистических методов, причем имеющих ряд некоторых преимуществ, то последним этапом предполагается проектирование, создание, обучение программного модуля на основе искусственной нейронной сети и проведение на его основе эксперимента по идентификации эмоций и их искренности.

1. Нахождение отличий в классификаторах для записей разных категорий

С помощью разработанной программы (рис. 2) на основе дискретного преобразования Фурье и автоматизации нахождения локальных максимумов амплитуд вычисляются соответствующие отношения частот.

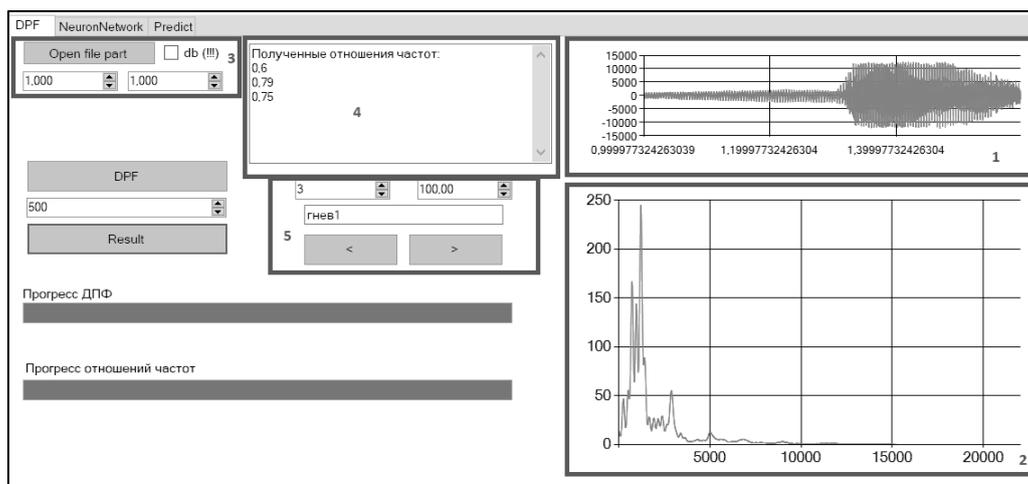


Рис. 2. Интерфейс программы для нахождения отношений частот для локальных максимумов амплитуд

Каждая запись была разделена на n_i отрезков в зависимости от длины записи, $i=1,2,..5$. Длина каждого отрезка составляла 1 сек, так как в работе [9] показано, что отношения, соответствующие музыкальным интервалам, лучше всего находятся при длительности интервала 1 сек.

В каждом отрезке определялись 7 частот v_{ij} ($i \leq n, 1 \leq j \leq 7$) с амплитудами локальных максимумов. В результате для каждой записи получались $n_i * 7$ измерений.

Для понимания различий между классификаторами, для записей различных категорий были построены диаграммы с наиболее часто встречающимися отношениями частот локальных максимумов. Полученные соотношения были сопоставлены с известными музыкальными интервалами. Абсолютная погрешность попадания в музыкальный интервал составляет не более 0,005, что является следствием математического округления вещественного числа до 2 знаков после запятой. В расчетах учитывались отношения меньше единицы и больше 0,5 (в рамках одной октавы).

Так, на рис. 3 можно увидеть диаграмму с наиболее часто встречаемыми отношениями частот локальных максимумов.



Рис. 3. Отношения частот для "истинной" радости

При выборе трех локальных максимумов наиболее часто встречаемым отношением [4] было 0,84 – большая терция, однако при выборе семи отношений, больше стали проявлять себя второстепенные отношения, но как раз именно они хорошо показывают различия между "истинными" и "фальшивыми эмоциями". Видно, что при таком подходе для искренней радости характерны: секста большая, чистая кварта, чистая квинта.

При сравнении с диаграммой ложной радости (рис. 4) можно заметить, что вместо большой сексты, проявляется малая секста, остальные музыкальные интервалы идентичны истинной эмоции.

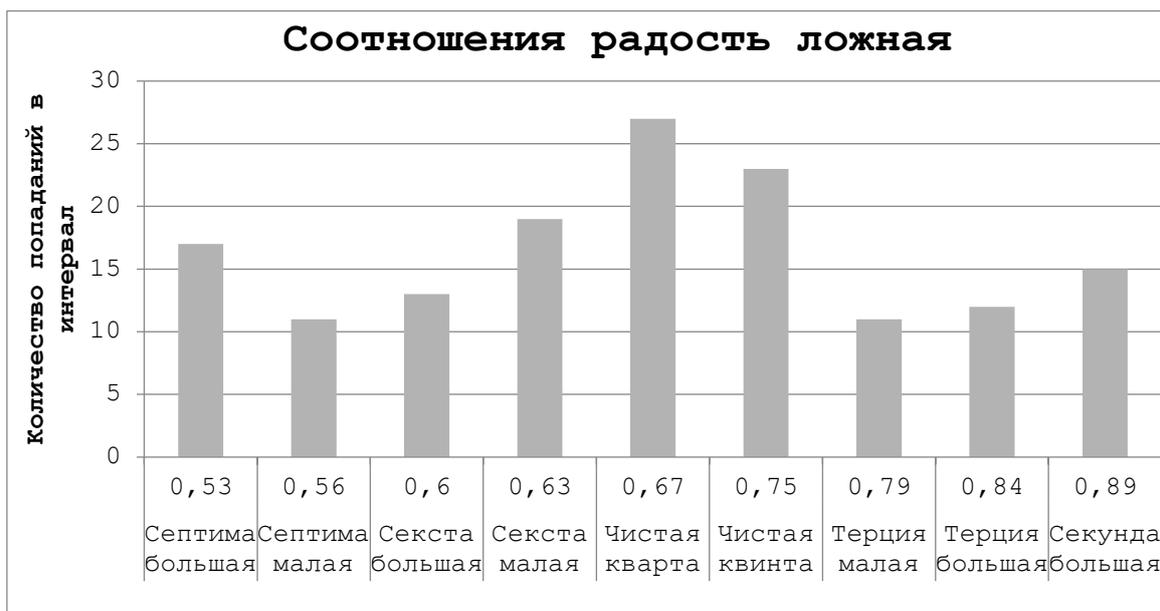


Рис. 4. Отношения частот для "фальшивой" радости

Если проанализировать изменения, с учетом того, что:

- большая секста – светлый, веселый, мягкий консонанс,
- малая секста – мягкий, нежный, грустный консонанс,

то здесь очевидны изменения в сторону грустного и мягкого звучания.

На рис. 5. представлены отношения частот локальных максимумов, характерных для записей с грустным звучанием.

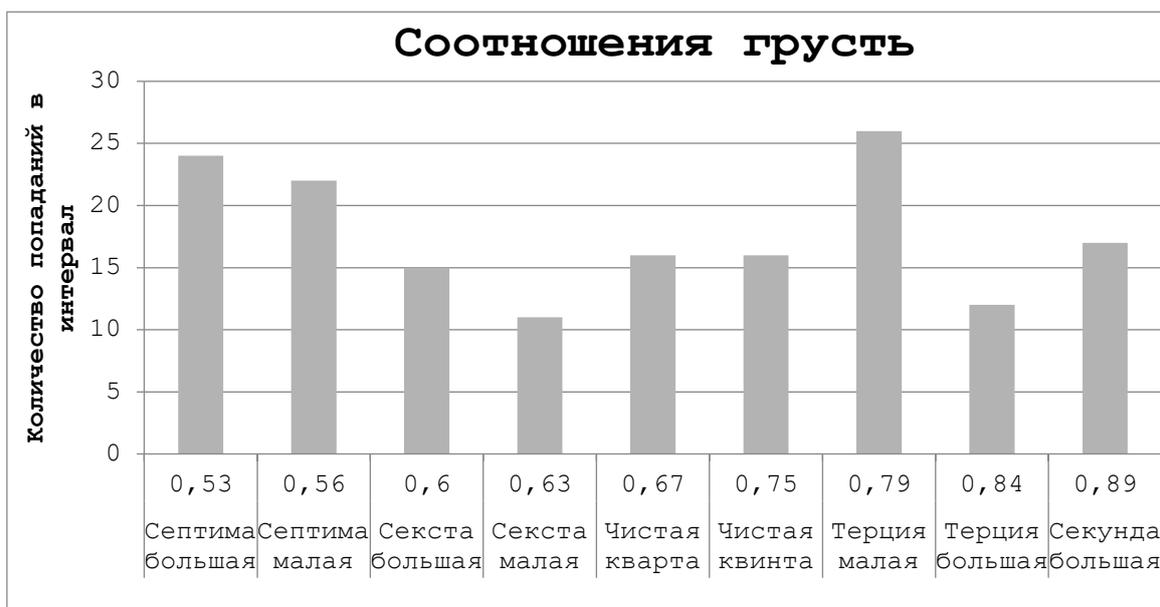


Рис. 5. Отношения частот для "истинной" грусти

А на рис. 6 можно увидеть существенные отличия.

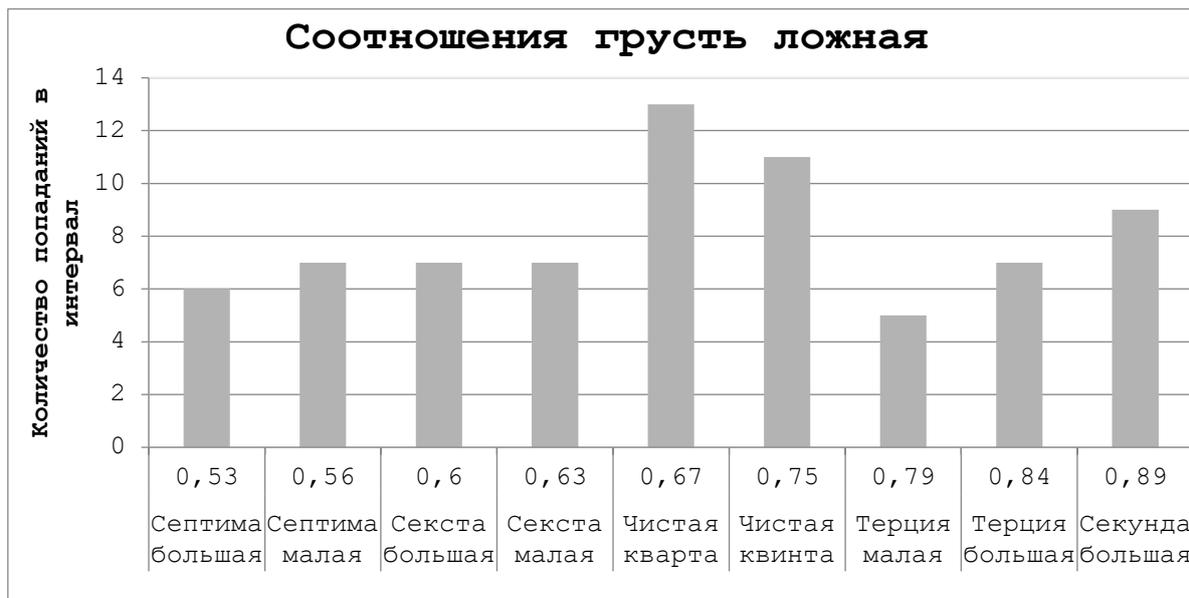


Рис. 6. Отношения частот для "фальшивой" грусти

Здесь видно, что отношение 0,79 – малая терция, характерное для грусти, перестает быть не только самым встречаемым, но и вообще становится самым наименее встречаемым из представленных. Для ложной грусти наиболее встречаемым оказалось отношение 0,67 – гордый, уверенный, твердый, яркий консонанс.

Продолжим анализировать и сравнивать рис. 7 для искреннего и рис. 8 для ложного гнева.

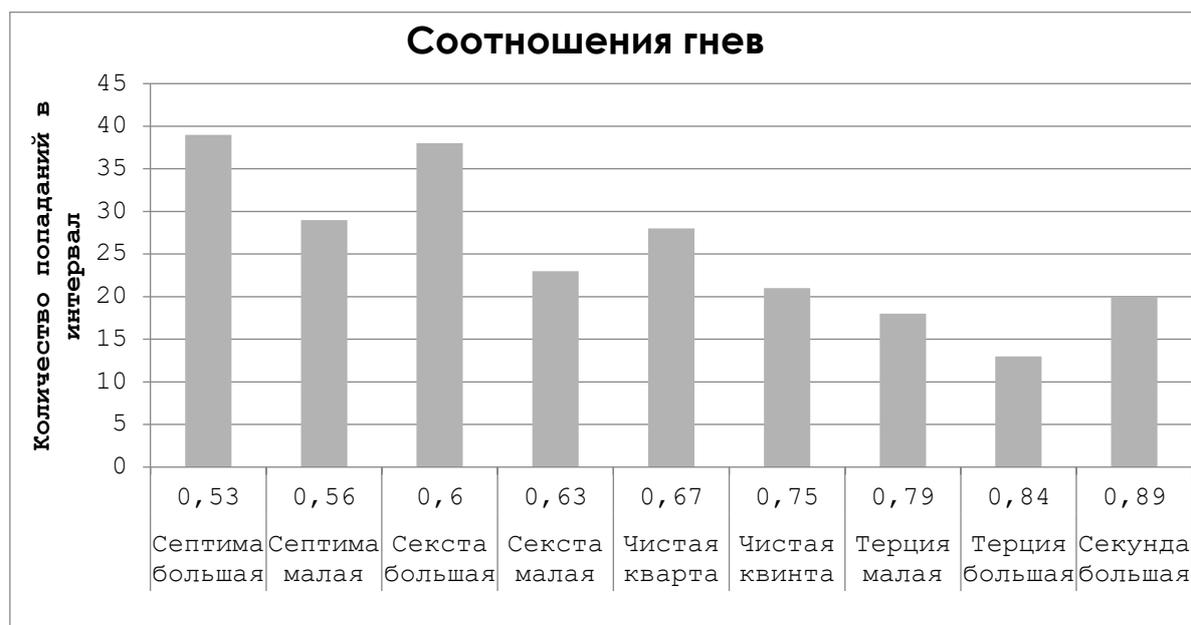


Рис. 7. Отношения частот для "истинного" гнева



Рис. 8. Отношения частот для "фальшивого" гнева

Можно заметить существенные отличия. В искреннем гневе преобладают большая септима и большая секста – резкий, острый диссонанс и светлый, веселый, мягкий консонанс. В ложном гневе уже преобладает чистая кварта – гордый, уверенный, твердый, яркий консонанс. Очевидно, что при ложном гневе человек себя лучше контролирует.

И последняя категория эмоций – страх. На рис. 9 представлены данные для искренней эмоции страха.

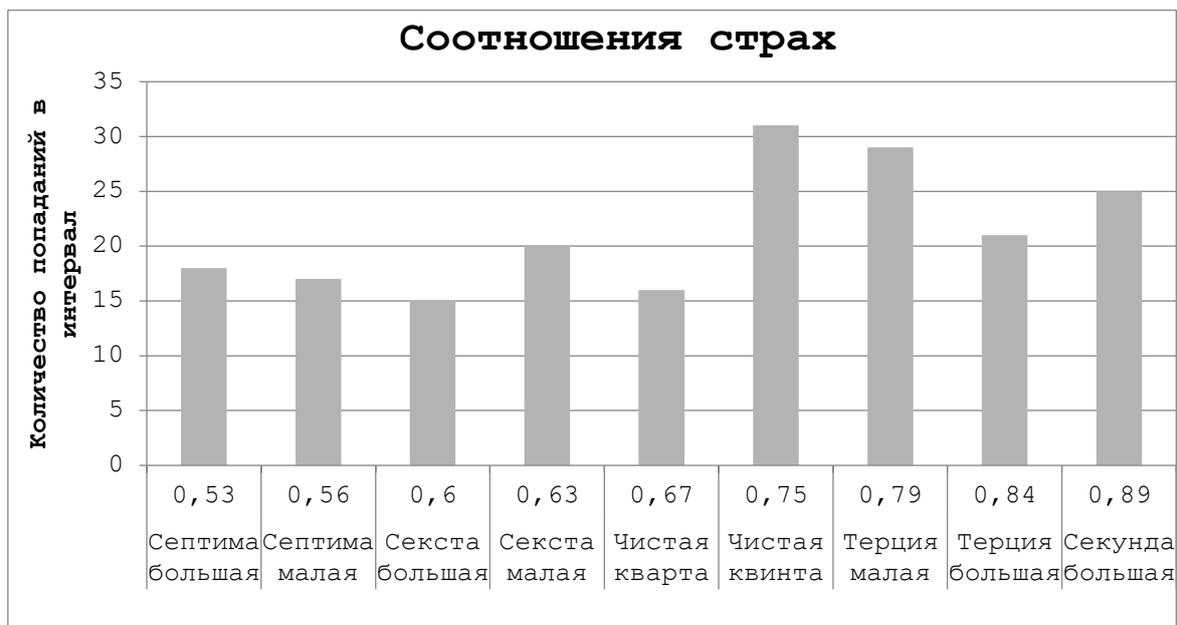


Рис. 9. Отношения частот для "истинного" страха

На рис. 10 так же, как и в предыдущих случаях, видны существенные отличия.

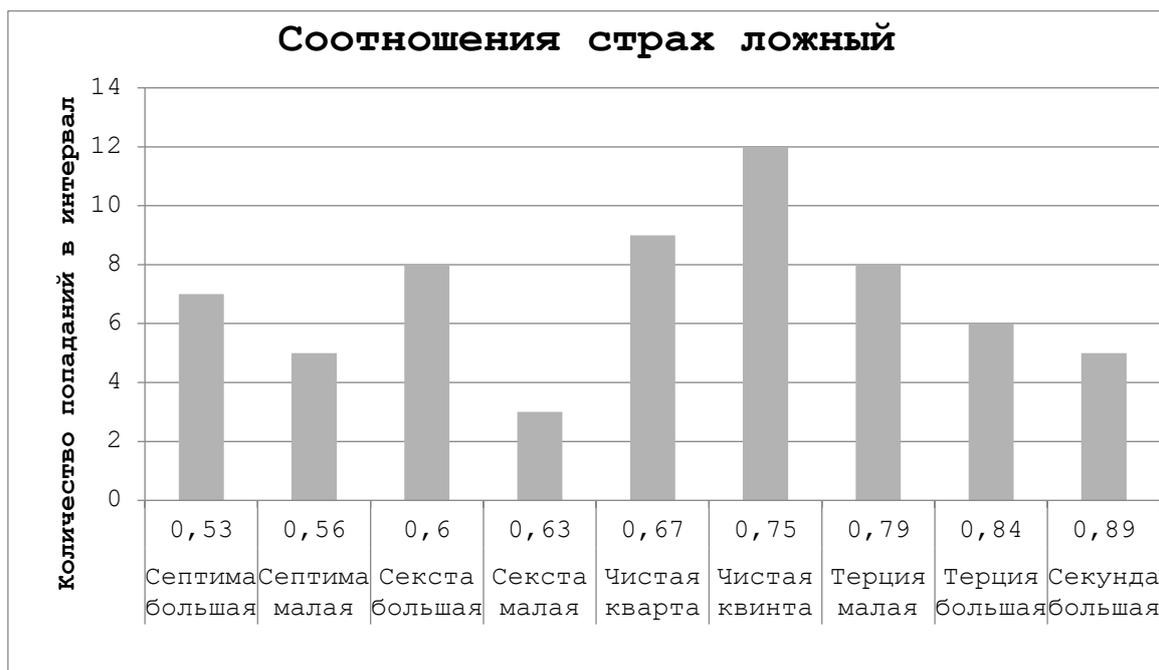


Рис. 10. Отношения частот для "истинного" страха

Сравнения диаграмм показывают, что во втором случае, как и во всех предыдущих случаях, существенное значение приобретает отношение 0,67 – чистая кварта – гордый, уверенный, твердый, яркий консонанс.

2. Создание и обучение искусственной нейронной сети

Как показано в предыдущем пункте, с помощью "ручного" сравнения статистических данных для отношений частот амплитуд локальных максимумов можно увидеть существенные отличия, как для различных категорий эмоций, так и для степени их искренности. Для автоматизации процесса идентификации эмоционального посыла говорящего был применен метод искусственной нейронной сети.

На основе предыдущего анализа был выбран набор из 10 входных параметров, который представляет собой "One-Hot Encoding" вектор, где каждый параметр может принимать значение либо 1, либо 0, в зависимости от наличия частоты определенного музыкального интервала:

- X1 – септима большая (0,53),
- X2 – септима малая (0,56),
- X3 – секста большая (0,6),
- X4 – секста малая (0,63),
- X5 – чистая кварта (0,67),
- X6 – чистая квинта (0,75),
- X7 – терция малая (0,79),
- X8 – терция большая (0,84),
- X9 – секунда большая (0,89),
- X10 – октава (0,5).

Выходные параметры также задаются вектором, в котором 8 элементов, по элементу на каждую эмоцию. Вектор содержит одну единицу у элемента, который определяет эмоцию записи, остальные значения равны нулю:

- D1 – Гнев,
- D2 – Страх,
- D3 – Грусть,
- D4 – Гнев (ложь),
- D5 – Грусть (ложь),
- D6 – Радость (ложь),
- D7 – Страх (ложь),
- D8 – Радость.

Исходя из результата предыдущего этапа, количество входных нейронов персептрона $N_x = 10$, количество нейронов выходного слоя $N_y = 8$.

Для определения объема обучающего множества существует эмпирическая формула $Q = 7 N_x + 15 = 85$, однако на практике обычно бывает достаточно и меньшего количества примеров. В нашем же случае примеров для обучения больше, чем значение, высчитанное по формуле $Q = 240$. Множество примеров следует разбить на обучающее L и тестирующее T , где в тестирующем множестве должно содержаться около 10–15 % от общего количества примеров. В данном случае это составит $L = 240$ и $T = 36$ соответственно.

По теореме Арнольда–Колмогорова–Хехт-Нильсена в качестве универсальной нейронной сети принято считать двухслойный персептрон с одним скрытым слоем, активационные функции у которого являются сигмоидными. Поэтому нейронная сеть будет иметь один скрытый слой. Необходимое количество синаптических связей можно высчитать по формуле

$$\frac{N_y \cdot Q}{1 + \log_2(Q)} \leq N_w \leq N_y \left(\frac{Q}{N_x} + 1 \right) (N_x + N_y + 1) + N_y$$

Таким образом $215,56 \leq N_w \leq 3808$.

Исходя из полученных выше данных, можно сосчитать количество нейронов в скрытом слое по формуле

$$N = \frac{N_w}{N_x + N_y}$$

В результате N варьируется от 11 до 211, поэтому будет выбрано то количество, где ошибка тестирования будет минимальна.

В результате оптимизации спроектированная нейронная сеть представляет собой персептрон, который имеет 10 входных нейронов, 11 нейронов на скрытом слое и 8 выходных нейронов (рис. 11).

В качестве алгоритма обучения использовался алгоритм упругого распространения.

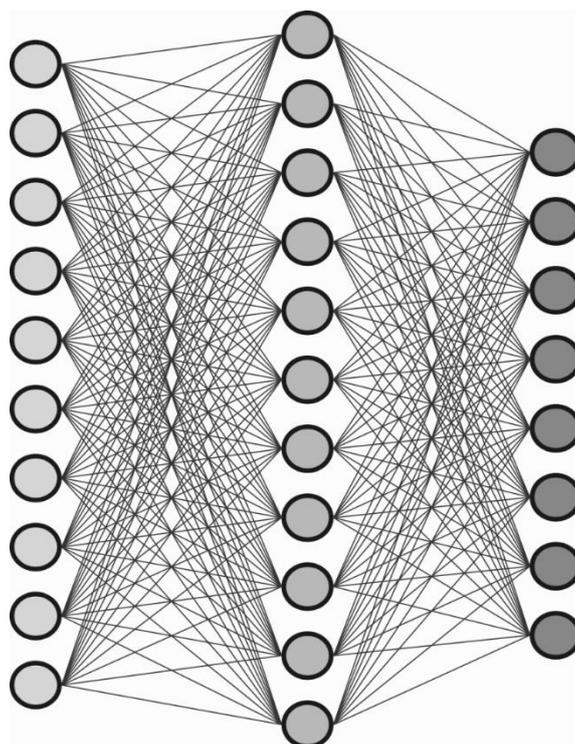


Рис. 11. Модель искусственной нейронной сети

Обучение искусственной нейронной сети производилось на основе 240 записей, по 30 записей на каждую категорию эмоции с учетом ее искренности.

После обучения был проведен эксперимент с целью определения возможности или невозможности использования метода отношения частот локальных максимумов для распознавания искренности эмоции. Для этого были выбраны по 5 дополнительных звуковых файлов на каждую группу эмоций, которые не присутствовали в обучающей и тестирующей выборке.

Результат тестирования представлен в табл. 2.

Таблица 2. Оценка количества корректно определенных эмоций

Эмоция	Эмоция определена верно	Эмоция определена не верно
Страх	4	1
Страх (ложная)	4	1
Радость	5	0
Радость(ложная)	3	2
Гнев	5	0
Гнев(ложная)	4	1
Грусть	5	0
Грусть(ложная)	4	1

Случаи, когда эмоция определялась верно или не верно, оценивались только с учетом ее истинности или ложности.

В итоге с использованием искусственной нейронной сети удалось разделить все 40 записей на соответствующие группы, при этом корректность распознавания эмоции и ее истинности составила 85 %.

Заключение

В результате исследований удалось показать, что метод, в котором в качестве классификаторов используются отношения частот локальных максимумов спектра звуковой волны, может быть хорошо применим к распознаванию искренности эмоции. Было показано, что в рассматриваемых категориях эмоций (радость, грусть, гнев, страх) в каждом случае между статистическими данными "истинной" и "фальшивой" эмоциями имеются существенные отличия. Общим же для всех категорий оказалось то, что в случае ложных эмоций существенную роль играет параметр отношения частот 0,67, соответствующий чистой кварте (гордый, уверенный, твердый, яркий консонанс). Данный консонанс показывает контроль говорящего человека над эмоцией.

Применение метода отношения частот локальных максимумов в искусственной нейронной сети показало хорошие результаты. Эмоции тестового множества (с учетом их истинности) были идентифицированы в 85 % случаев.

Список источников

1. *Российские ученые создали RiskControl, который распознает ложь по голосу.* <https://www.computerra.ru> – журнал "Компьютерра". 2021. [Электронный ресурс]. Режим доступа: <https://www.computerra.ru/275450/rossijskie-uchenye-sozdali-riskcontrol-kotoryj-raspoznat-lozh-po-golosu/> (дата обращения: 19.11.2024).
2. *Пеленицын А.Б., Степанов А.А.* Что надо знать о так называемых "детекторах лжи по голосу". [Электронный ресурс] / А.Б. Пеленицын, А.А. Степанов. Электрон. текстовые дан. Режим доступа: http://poligraf.sp.ru/stati/abpelenicyn_aastepanov-detektoru_lzhi_po_golosu.html, свободный (дата обращения: 19.11.2024).
3. *Изард К.Э.* Психология эмоций / пер. с англ. СПб: Изд-во "Питер", 1999. 464 с.
4. *Shkaraputa, A., Kolcherina, A., Mishlanova, M.* (2022). Determining of the Emotional State of a Person Using Computer Analysis of Sound Wave Parameters. In: Rocha, A., Isaeva, E. (eds) Science and Global Challenges of the 21st Century - Science and Technology. Perm Forum 2021. Lecture Notes in Networks and Systems, Vol. 342. Springer, Cham. https://doi.org/10.1007/978-3-030-89477-1_6.
5. *Алимуратов А.К., Чураков П.П.* Обзор и классификация методов обработки речевых сигналов в системах распознавания речи. Измерения. Мониторинг. Контроль. 2(12) 27-35 (2015).
6. *Xuedong H.* Spoken Language Processing: A Guide to Theory, Algorithm and System Development. Huang Xuedong. New Jersey. Prentice Hall PTR (2001).
7. *Mehmet Cenk Sezgin, Bilge Günsel & Gunes Karabulut Kurt.* Perceptual audio features for emotion detection. EURASIP Journal on Audio, Speech, and Music Processing volume 2012, Article number: 16 (2012) <https://doi.org/10.1186/1687-4722-2012-16>.
8. *Рахманенко И.А., Мецераков Р.В.* Анализ идентификационных признаков в речевых данных с помощью GMM-UBM верификации диктора. С-Пб.: Тр. СПИИРАН, 2017, С. 32–50.
9. *Близнюк О.А., Шкарапута А.П.* Исследование применения метода определения базовых эмоций на основе отношения частот спектра звуковой волны // Вестник Пермского университета. Математика. Механика. Информатика. 2017. Вып. 4(39). С. 86–91.

References

1. <https://www.computerra.ru> - Computerra Magazine. 2021. [Electronic resource]. Mode of access: <https://www.computerra.ru/275450/rossijskie-uchenye-sozdali-riskcontrol-kotoryj-raspoznat-lozh-po-golosu/> (date of address: 19.11.2024).

2. Pelenitsyn, A.B., Stepanov, A.A. WHAT TO KNOW ABOUT THE SO-called "VOICE LIE DETECTORS" [Electronic resource] / ePelenitsyn A.B., A.A.Stepanov - Electronic text data [Electronic resource] / Pelenitsyn A.B. Pelenitsyn, A.A.Stepanov - Electronic text dan. Mode of access: http://poligraf.sp.ru/stati/abpelenicyn_aastepanov_detektory_lzhi_po_golosu.html, free (date of reference: 19.11.2024).
3. Izard, K.E. The Psychology of Emotions. St. Petersburg. Peter Publishers (2008).
4. Shkaraputa, A., Kolcherina, A., Mishlanova, M. (2022). Determining of the Emotional State of a Person Using Computer Analysis of Sound Wave Parameters. In: Rocha, A., Isaeva, E. (eds) Science and Global Challenges of the 21st Century - Science and Technology. Perm Forum 2021. Lecture Notes in Networks and Systems, vol 342. Springer, Cham. https://doi.org/10.1007/978-3-030-89477-1_6.
5. Alimuradov, A. K., Churakov, P. P. (2015). Review and classification of speech signal processing methods in speech recognition systems. Measurement. Monitoring. Control. 2(12) 27-35.
6. Xuedong, H. Spoken Language Processing: A Guide to Theory, Algorithm and System Development. Huang Xuedong. New Jersey. Prentice Hall PTR (2001).
7. Mehmet Cenk Sezgin, Bilge Günsel & Gunes Karabulut Kurt. Perceptual audio features for emotion detection. EURASIP Journal on Audio, Speech, and Music Processing volume 2012, Article number: 16 (2012). <https://doi.org/10.1186/1687-4722-2012-16>.
8. Rakhmanenko, I.A., Meshcheryakov, R.V. (2017). Analysis of identification features in speech data using GMM-UBM speaker verification. Proceedings of SPIIRAS, pp. 32-50. St. Petersburg.
9. Bliznyuk, O. A., Shkaraputa, A. P. (2027). The study of the application of the method of determining the basic emotions based on the ratio of frequencies of the sound wave spectrum. Bulletin of Perm University. Mathematics. Mechanics. Informatics. 4(39), pp. 86-91.

Информация об авторах:

И. В. Быков – бакалавр по направлению "Прикладная математика и информатика" механико-математического факультета Пермского государственного национального исследовательского университета (614068, Россия, г. Пермь, ул. Букирева, 15);

Н. Ю. Ротанева – кандидат педагогических наук, доцент, заведующий кафедрой системного анализа и информационных технологий ФГБОУ ВО "Мариупольский государственный университет имени А.И. Куинджи", AuthorID: 1191430;

А. П. Шкарапута – кандидат физико-математических наук, доцент института компьютерных наук и технологий Пермского государственного национального исследовательского университета (614068, Россия, г. Пермь, ул. Букирева, 15), AuthorID: 69440.

Information about the authors:

I. V. Bykov – Bachelor's degree in Applied Mathematics and Informatics, Faculty of Mechanics and Mathematics, Perm State University (15, Bukireva St., Perm, Russia, 614068);

N. Yu. Rotaneva – Candidate of Pedagogical Sciences, Associate Professor, Head of the Department of System Analysis and Information Technologies of the Mariupol State University named after A.I. Kuindzhi, AuthorID: 1191430;

A. P. Shkaraputa – Candidate of Physical and Mathematical Sciences, Associate Professor of the Institute of Computer Science and Technology of Perm State University (15, Bukireva St., Perm, Russia, 614068), AuthorID: 69440.