

Профессор Г.В. Абрамов, доцент Л.А. Коробова,
доцент А.Л. Ивашин, аспирант И.А. Матыцина

(Воронеж. гос. ун-т. инж. технол.) кафедра информационных технологий, моделирования и управления, тел. (473) 255-25-50

E-mail: irina210390@mail.ru

Professor G.V. Abramov, assistant Professor L.A. Korobova,
assistant professor A.L. Ivashin, graduate I.A. Matytsina

(Voronezh state university of engineering technologies)

Department of information technology, modeling and management, phone (473) 255-25-50

E-mail: irina210390@mail.ru

Анализ и использование математических методов для распознавания звуковых сигналов

The analysis and use of mathematical methods for the detection of sound signals

Реферат. Распознавание голоса – широко изучаемая и изученная в науке тема, а распознавание звуковых сигналов, в том числе кашлей пациентов в длительной звуковой записи – практически не изученная. Для распознавания звуковых сигналов и их подсчета были проанализированы и использованы математические методы, такие как корреляционный анализ, быстрое преобразование Фурье, нечеткая логика. Рассматривая каждый метод по очереди, подбирались параметры дающие наилучшие результаты распознавания при использовании того или иного метода. Корреляционный анализ позволяет сделать выводы о количестве кашлевых моментов для выбранного эталона. Быстрое преобразование Фурье позволило выделить такие частотные диапазоны, в которые попадают только кашли и не попадают шумы, экспериментальным путем было выявлено 6 таких диапазонов. Использование нечеткой логики улучшило анализ звуковой записи и позволило производить выбор кашлевых моментов с большей степенью уверенности, но однозначности результатов выбора достичь так и не удалось. Использовалась нечеткая логика, позволяющая расширить границы распознавания, позволяя градировать результат на кашель, возможно кашель и шум. При анализе полученных данных по каждому математическому методу пришли к выводу, что по отдельности их использование невозможно. Для достижения поставленной цели необходимо составить комплекс математических правил вывода, которые позволят распознавать звуковые сигналы с более высокой точностью. Рассмотрев спектрограмму звукового фрагмента, выделить зоны, в которых необходимо производить анализ для выявления похожести звуковых записей. Выделенные зоны кашля имеют четкие значения по каждой из координат спектрограммы. Это позволило в дальнейшем при синтезировании правил вывода использовать точные значения для границ параметров этих правил. Под полученный комплекс правил модифицировали разработанный программный продукт. Протестировали запись и получили результаты.

Summary. Voice recognition is a widely studied and explored in science, and the recognition of audio signals, including cough patients in long audio recording is practically not studied. To recognize the audio signals and their counting were analyzed and used mathematical methods such as correlation analysis, fast Fourier transform, fuzzy logic. Looking at each method in turn, were selected for the parameters gives the best recognition results when using either method. Correlation analysis allows to draw conclusions about the number of cough moments for the selected standard. Fast Fourier transform allowed to allocate these frequency ranges, which are the only cough and do not get the noise was experimentally identified 6 of these ranges. The use of fuzzy logic has improved the analysis of sound recordings and allowed to make the choice of the cough moments with more certainty, but certainty of results to achieve and failed. Used fuzzy logic, which allows to expand the boundaries of recognition, allowing graded result in coughing may cough and noise. When analyzing the data obtained for each of the mathematical method, came to the conclusion that individually cannot be used. To achieve this goal it is necessary to compile a set of mathematical rules of inference that allow to recognize the sound signals with higher accuracy. Having considered the spectrogram of a sound clip, select the zone in which it is necessary to conduct an analysis to identify the similarity of sound recordings. The zones of cough have distinct values for each of the coordinates of the spectrogram. This allowed later when synthesizing inference rules use the exact values for the parameters of these rules. Under the resulting set of rules modified the developed software. Tested and received results.

Ключевые слова: звуковые сигналы, анализ, корреляция, нечеткая логика, правила вывода

Keywords: audible alarm, analysis, correlation, fuzzy logic, inference rules

В настоящее время, благодаря накоплению специализированных знаний и опыта, математическое моделирование стало мощным инструментом анализа процессов для распознавания звуковых сигналов, а в частности - кашлевых моментов в звуковых записях. В особой мере это касается изучения и создания комплекса правил вывода для воздействия на разнообразные факторы и характеристики звуковой записи. Возможности прямой обработки записи при этом, как правило, ограничены, в то время как созданная с использованием

доступных экспериментальных данных математическая модель позволяет обеспечить подробную детализацию и оценить взаимное влияние различных параметров друг на друга, а также на распознавание звуковых сигналов в целом. Важнейшим условием эффективности вычислительного эксперимента является адекватность математической модели протекающим в распознавании кашлевых моментов.

© Абрамов Г.В., Коробова Л.А.,
Ивашин А.Л., Матыцина И.А., 2015

Математическая модель представляет собой разработку правил вывода, основанных на нечеткой логике. В качестве параметров правил вывода приняты реальные математические понятия, такие как коэффициент корреляции Пирсона, среднее квадратическое отклонение, дисперсия, Байесовская оценка, интервальная оценка, быстрое преобразование Фурье, интегральная оценка.

1. Корреляционный анализ. Данный термин означает совокупность основанных на математической теории корреляции методов обнаружения корреляционной зависимости между двумя случайными признаками или факторами. Корреляционный анализ экспериментальных данных включает в себя следующие основные практические приемы:

1) вычисление коэффициента корреляции Пирсона по формуле (1):

$$r_{xy}^P = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (1)$$

2) проверка статистической гипотезы значимости силы корреляционной связи: - сильная: $\pm 0,7$ до ± 1 ; - средняя: $\pm 0,3$ до $\pm 0,699$; - слабая: 0 до $\pm 0,299$ [1];

3) построение корреляционного поля и составление корреляционной таблицы по данным рассчитанным по двум коэффициентам корреляции Пирсона и Спирмена.

Таблица 1

Анализ данных по коэффициенту корреляции Пирсона и Спирмена

	Коэффициент корреляции Пирсона	Ранговый коэффициент корреляции Спирмена
Кашель эталонный	1	1
min значение коэффициента корреляции кашля пациента	0,12	0,99995
max значение коэффициента корреляции кашля пациента	0,24	0,99998
max значение коэффициента шума	0,22	0,99997

Первоначально для распознавания кашлевых моментов и шумовых фрагментов использовался только один коэффициент корреляции Пирсона. Алгоритм распознавания был реализован на языке программирования Delphi. Результаты апробации тестовой записи представлены на рисунке 1.

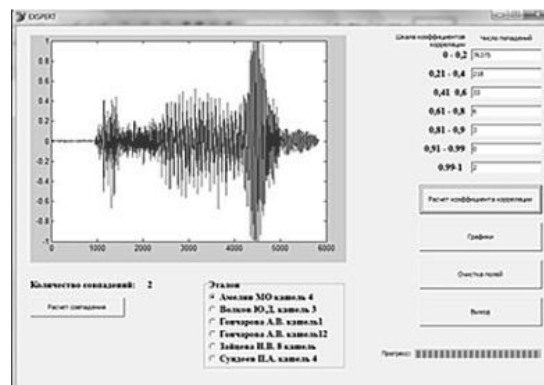


Рисунок 1. Программный продукт, реализованный в Delphi

В ячейках верхней правой части экранной формы отображается количество просчитанных коэффициентов корреляции Пирсона, разбитых по группам от 0 до 0,2; от 0,2 до 0,4; от 0,4 до 0,6; от 0,6 до 0,8; от 0,8 до 0,9; от 0,9 до 0,99 и от 0,99 до 1. Нажатие кнопки «графики» (рисунок 1) дает возможность просмотреть график с рассчитанными максимумами коэффициентов корреляции (рисунок 2).

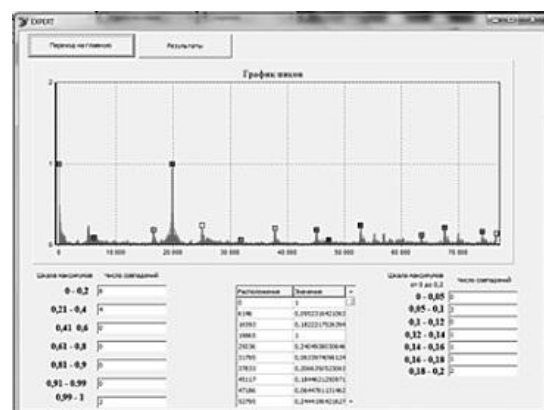


Рисунок 2. График с максимумами коэффициентов корреляции

На графике отображаются всплески похожести звуковых сигналов на эталон, чем ближе к единице, тем степень похожести больше. По результату можно сделать вывод о количестве кашлевых моментов для выбранного эталона. Использование только одного параметра (коэффициента корреляции) для определения кашлевых моментов в звуковой записи дало низкую точность. Анализ результатов показал, что нельзя использовать только метод корреляционного анализа, т.к. максимальное значение коэффициента корреляции шума больше минимального значения коэффициента корреляции кашля пациента (таблица 1).

Для повышения точности и надежности результатов при распознавании кашлевых моментов следующим шагом стало применение быстрого преобразования Фурье. Данная проце-

дура была осуществлена в программном пакете Maple с помощью встроенной функции $\text{fft}(f)$. Так, на рисунке 3 показано, что спектры кашля и шума в большинстве случаев различаются [2].

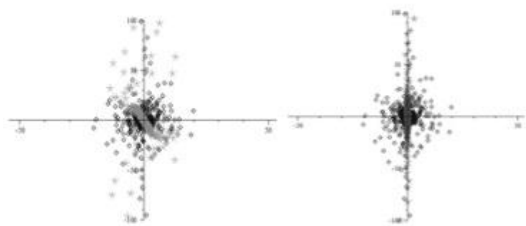


Рисунок 3. Изображение кашля и шума: $\diamond$ - кашель; * - шум

Преобразованный по быстрому преобразованию Фурье сигнал близок к прямой линии относительно оси ординат или под углом к ней. Однако существуют шумы очень похожие на кашель (рисунок 4). Отсюда следует, что на данный подход не дает возможность делать однозначные выводы.

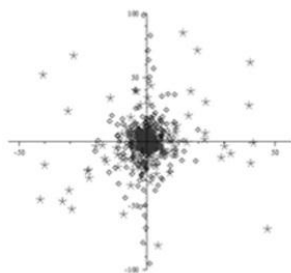


Рисунок 4. Изображение кашля и шума близкого к кашлю: $\diamond$ - кашель; * - шум

Следующий этап обработки звукового сигнала заключается в необходимости выделения таких частотных диапазонов, в которые

будут попадать только кашли и не попадут шумы. Экспериментальным путем, анализируя графики кашлей и шумов, такая процедура была проведена и выделено 6 диапазонов.

Таким образом, каждый шаг при выделении кашлевых моментов из многовременной записи вносил свои неточности и неопределенности в общий алгоритм. Поэтому было принято решение по использованию механизмов нечеткой логики, который позволит объединить данные по исследованию коэффициентов корреляции и быстрого преобразования Фурье.

2. Нечеткая логика. Лингвистические переменные предназначены в основном для анализа сложных или плохо определенных явлений. Использование словесных описаний делает возможным анализ систем настолько сложных, что они недоступны обычному математическому анализу. Более точно структура лингвистической переменной описывается набором (N, T, X, G, M) , в котором N - название этой переменной; T - терм-множество N , т.е. совокупность ее лингвистических значений; X - универсальное множество с базовой переменной x ; G - синтаксическое правило, которое может быть задано в форме бесконтекстной грамматики, порождающей термы множества T ; M - семантическое правило, которое каждому лингвистическому значению t ставит в соответствие его смысл $M(t)$, причем $M(t)$ обозначает нечеткое подмножество множества X [2].

По показателям, представленным выше, введем 4 лингвистические переменные: коэффициент корреляции, отклонение амплитуд, среднее значение положительной амплитуды и среднее значение отрицательной амплитуды (таблица 2).

Таблица 2

Описание лингвистических переменных

Переменная из набора	Коэффициент корреляции	Отклонение амплитуд	Среднее значение отрицательной амплитуды	Среднее значение положительной амплитуды
N	Коэффициент корреляции	Отклонение амплитуд	Среднее значение отрицательной амплитуды	Среднее значение положительной амплитуды
T	Высокий, средний, низкий	Высокое, среднее, низкое	Высокое, среднее, низкое	Высокое, среднее, низкое
X	[0;1]	[0;1]	[0;1]	[0;1]
G	Не, очень, не очень	Не, очень, не очень	Не, очень, не очень	Не, очень, не очень

В результате сформируем следующие правила нечеткого вывода [3]:

ЕСЛИ все показатели низкие, ТО Шум;

ЕСЛИ один из показателей средний остальные низкие, ТО Шум;

ЕСЛИ один из показателей высокий остальные низкие, ТО Шум;

ЕСЛИ все показатели средний, ТО возможно кашель;

ЕСЛИ один из показателей низкий остальные высокие, ТО возможно кашель;

ЕСЛИ один из показателей высокий остальные средние, ТО возможно кашель;

ЕСЛИ одна половина показателей высокие, а другая низкие, ТО возможно кашель;

ЕСЛИ один показатель средний остальных высокие, ТО кашель;

ЕСЛИ все показатели высокие, ТО кашель.

Использование нечеткой логики улучшило анализ звуковой записи и позволило производить выбор кашлевых моментов с большей степенью уверенности, но однозначности результатов выбора достичь так и не удалось. К имеющимся параметрам анализа записи коэффициенту корреляции и преобразованию Фурье добавили дополнительные параметры интегральной ошибки, средней амплитуды и чувствительности настройки выбора. Алгоритм, использующий нечеткую логику, реализован на языке программирования Java SE 7. Результаты апробации тестовой записи представлены на рисунке 5.



Рисунок 5. Программный продукт с фиксированными показателями

В результате обработки выбранной записи происходит сортировка фрагментов на кашли, близкие к кашлям и шумы по определенному алгоритму. Итогом работы программы являются выделенные по цветам дорожки: - светло-зеленым цветом выделены звуковые дорожки, которые являются кашлями; - темно-зеленым цветом – возможно кашли; - серым – шумы.

В данной реализации предусмотрена возможность прослушивания любого звукового фрагмента. Для этого на дорожке интересующего фрагмента в левой части экрана необходимо кликнуть двойным щелчком левой клавиши мыши. Прослушав запись фрагмента можно переместить его в разряд кашлей или шумов. Для этого на дорожке интересующего фрагмента необходимо сделать один клик правой клавишей мыши и выбрать в контекстном меню вкладку «Установить как кашель» или «Уста-

новить как шум». Также можно изменить эталон – клик на вкладке «Установить как эталон».

Использованные параметры преобразования в правила, которые позволили распознать звуковые сигналы, но точность результатов не была высока, т.к. градация разделения звуковой записи проходила по 3 составляющим: «кашель», «возможно кашель» и «шум». Таким образом, появляется необходимость в модификации (расширении, дополнении и качественном изменении) алгоритма анализа звуковой записи и выбора кашлевых моментов. Следующим этапом является добавление в алгоритм некоторого механизма, который позволит применить для формализации помимо показателей, которые уже использовали при разработке правил вывода, еще и кластерный анализ.

3. Синтезирование правил вывода.

Обрабатываемый звуковой фрагмент представлен в звуковой форме и в виде спектрограммы. Прослушав звуковой фрагмент и полагаясь только на свои индивидуальные способности (слух) специалист может удостовериться, является ли данный фрагмент – кашлем или шумом. Просмотрев трехмерную спектрограмму в координатах «время – частота – амплитуда» и сравнив ее визуально со спектрограммой эталона, можно выделить зоны, в которых необходимо производить анализ для выявления похожеści звуковых записей. Выделенные зоны кашля теперь имеют четкие значения по каждой из координат спектрограммы. Это позволяет в дальнейшем при разработке нечетких правил вывода использовать точные значения для границ параметров этих правил [5].

Представим некоторые правила вывода, которые могут быть применимы к любому параметру спектрограммы:

- критерий корреляций $(p, j, r, k, l, m, q, i) > const1$;

- критерий суммы значений в окне на спектрограмме в сравнении с эталоном $(p, j, r, k, l, m, h) > const2$,

где p – выбранный эталон;

j – значение ширины статического окна при построение спектрограммы;

r, k – координаты левой и правой границы окна (по времени);

l, m – координаты нижней и верхней границы окна (по частоте);

q, i – координаты левой и правой границы окна (по времени) расчета коэффициента;

h – порог разницы;

$const1, const2$ – некоторое число от 0 до 1.

4. Визуализация модифицированного программного продукта.

Модификацией данного программного продукта является возможность самостоятельно вырабатывать комплексы правил вывода, что позволяет расширить возможность точного распознавания [4].

Работа программного продукта также заключается в распознавании кашлевых моментов из звуковой записи пациента (рисунок 6). Первоначально программа разбивает всю звуковую запись на отрезки равной величины. Выбирается эталон. Затем, используя правила вывода, производятся расчеты относительно выбранного эталона и по значениям выбран-

ным правилам определяется, является ли обрабатываемый фрагмент кашлем или нет. Если фрагмент – шум, то он становится красного цвета, если кашель – цвет остается прежним.

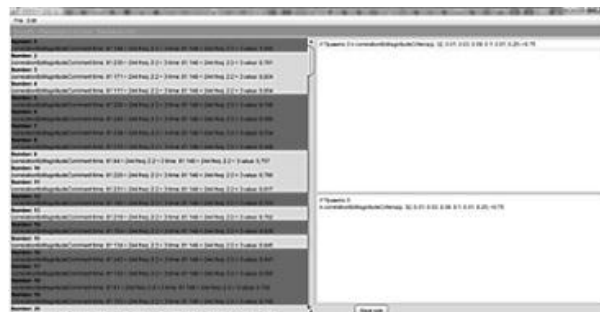


Рисунок 6. Модифицированный программный продукт

ЛИТЕРАТУРА

1 McLuckie A. Respiratory disease and its management. New York: Springer, 2009. 51 p.

2 Заде Л.А. Понятие лингвистической переменной и его применение к принятию приближенных решений. М: Мир, 1976. 165 с.

3 Абрамов Г.В., Овсянников Е.С., Коробова Л.А., Матыцина И.А. Разработка информационной системы для диагностики состояния больных легочными заболеваниями // Сборник трудов II международной научной интернет-конференции «Математическое и компьютерное моделирование в биологии и химии» в 2 томах. Казань: ИП Синяев Д.Н., 2013. Т.2. С. 4-10.

4 Абрамов Г.В., Коробова Л.А., Матыцина И.А. Разработка правил вывода при распознавании звуковых сигналов // IV международная научно-практическая конференция "Фундаментальная наука и технологии - перспективные разработки", USA. North Charleston, 2014. Т 3. С. 145-148.

5 Шилдт Г. Полный справочник по Java. Java SE™ 6 Edition (7-е издание). М.: Вильямс, 2007. С. 1040.

REFERENCES

1 McLuckie A. Respiratory disease and its management. New York: Springer, 2009. 51 p

2 Zadeh L. A. Ponyatie lingvisticheskoi peremennoi i ego primeneniye k prinytiyu priblizhennykh reshenii [Concept of a linguistic variable and its application explained to approximate reasoning]. Moscow, Mir, 1976. 165 p. (In Russ.).

3 Abramov G. V., Ovsyannikov E. S., Korobova L. A., Matytsin I. A. Development of information system for diagnostics of a condition of patients with pulmonary diseases. Sbornik trudov II mezhdunarodnoi nauchnoi internet-konferentsii "Marematicheskoe i komp'yuternoe modelirovanie v biologii i khimii" [Collection of labor in the II international scientific Internet-conference "Mathematical and computer modeling in biology and chemistry" in 2 volumes]. Kazan': Sinyayev D. N., 2013, vol. 2, pp. 4-10. (In Russ.).

4 Abramov G. V., Korobova L. A., Matytsin I. A. Development of rules of inference in recognition of audio signals. IV mezhdunarodnayanayachno-prakticheskaya konferentsiya "Fundamental'naya nauka i tekhnologii – perspektivnye razrabotki" [IV international scientific-practical conference "Fundamental science and technology - promising developments"]. North Charleston, 2014, vol. 3, pp. 145-148. (In Russ.).

5 Shildt G. Polnyi spravochnik po Java. Java™ SE 6 Edition [Complete reference to Java. Java™ SE 6 Edition (7th edition)]. Moscow, Williams, 2007. 1040 p. (In Russ.).