# ПРОГРАММНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС АНАЛИЗА ЭМОЦИОНАЛЬНО ОКРАШЕННОЙ РЕЧИ

**А. В. Ткаченя, С. Г. Мулярчик**

*Белорусский государственный университет*
*Минск, Беларусь*
*e-mail: tkachenia@gmail.com, mulyarchik@bsu.by*

Рассматривается проблематика создания программно-методического комплекса с открытым интерфейсом программирования приложений в области анализа речевых сигналов. Описывается разработанная на основе реализованного комплекса система, которая позволяет быстро расширять свою функциональность, предназначенная для анализа эмоциональной речи, изучения методов автоматического распознавания эмоциональной слитной речи и влияния эмоциональной окраски речевого сигнала на точность распознавания.

*Ключевые слова*: анализ речевых сигналов; распознавание эмоционально окрашенной речи; классификация эмоций.

# SOFTWARE-METHODICAL FRAMEWORK FOR EMOTIONAL COLORED SPEECH ANALYSIS

**A. V. Tkachenia, S. G. Mulyarchik**

*Belarus State University*
*Minsk, Belarus*

In the paper problems of designing software-methodical framework with an open application programming interface for speech signals analysis are studied. It describes developed system for emotional speech analysis, automatic emotional speech recognition and influence of emotional coloring on speech recognition accuracy, which is based on the released framework and provides the ability to quickly expand its functionality.

*Keywords*: speech signals analysis; emotional colored speech recognition; emotion classification.

## ВВЕДЕНИЕ

Задача анализа речевых сигналов широко востребована в современном мире. Развитие информационного общества постоянно предъявляет все больше требований по совершенствованию автоматических подходов к анализу речи и их применению для решения прикладных задач, количество которых увеличивается с каждым годом. Однако бурное развитие наукоемких алгоритмов голосового анализа приводит к возникновению определенных трудностей не только с изучением принципов их работы, но и в практическом их использовании. В частности одной из немаловажных проблем

является информативное и понятное представление полученных результатов в виде графических или числовых данных.

С целью обеспечить доступ к современным технологиям решения научно-практических задач, а также сблизить научную и практическую составляющую в подготовке специалистов, был разработан программно-методический комплекс, предоставляющий широкий спектр возможностей для проведения экспериментов и исследований в сфере анализа речевых сигналов.

## ФУНКЦИОНАЛЬНОСТЬ ПРОГРАММНО-МЕТОДИЧЕСКОГО КОМПЛЕКСА ДЛЯ АНАЛИЗА РЕЧЕВЫХ СИГНАЛОВ

Структурно разработанный программно-методический комплекс можно разделить на две составляющие: функциональная часть и графический интерфейс. Для решения разнообразного круга задач функциональность разработанного комплекса предоставляет как общие возможности:

чтение и сохранение многоканальных аудио файлов (WAV и MP3);

многопоточная обработка речевых баз данных;

генерация файлов с отчетами о полученных результатах,

так и специализированные алгоритмы анализа речевых сигналов:

полосовая фильтрация аудио сигналов [1];

процедура линейного предсказания [2];

быстрое преобразование Фурье [3];

дискретно-косинусное преобразование [4];

вейвлет-преобразование [5];

применение частотных шкал (например, шкала барков или мел-частот [6]);

расчет просодических параметров речи [7];

вычисление частоты основного тона (ЧОТ) [8];

метод опорных векторов (МОВ) [9];

алгоритм динамической трансформации шкалы времени (ДТВ) [10];

скрытые марковские модели (СММ) [11];

генерация сетей спутывания и выделение триггерных пар слов [12],

при этом библиотека поддерживаемых алгоритмов может легко пополняться по мере необходимости. В тоже время на основе комбинации имеющихся алгоритмов можно получать как разнообразные вектора признаков (например, мел-частотные кепстральные коэффициентов [13] или перцептивные коэффициенты линейного предсказания [14]), так и различные методики анализа речи (например, классификация эмоций на основе МОВ [15] или верификация слов на основе ДТВ [16]).

Программно-методический комплекс создан в свободной кроссплатформенной среде разработки Code::Blocks при использовании компилятора MinGW/GCC версии 4.9.1 на языке программирования C++. Графический интерфейс реализован при помощи кроссплатформенной библиотеки Qt версии 5.4.0. Обновление графических данных происходит в реальном масштабе времени и не требует больших вычислительных ресурсов. Графический интерфейс предоставляет следующие возможности:

воспроизведение и запись аудио данных;

отображение гистограмм;

круговые пиктограммы;

отображение графиков, с возможностью выделения временных участков, а также подсветкой отдельных рядов данных и их участков различным цветом.

Использование языка программирования C++ и кроссплатформенной библиотеки Qt позволяет добиться наилучшей совместимости с различными операционными системами и обеспечить достаточную производительность необходимую для работы в реальном масштабе времени, а также допускает расширение и модификацию возможностей программно-методического комплекса за счет открытости исходных кодов. Реализованная функциональная и графическая часть позволяет решать широкий спектр задач в сфере анализа речевых сигналов и быстро создавать прикладные программы, которые могут быть использованы в экспертных организациях и учреждениях, а также в исследовательской работе или в учебных целях.

## СИСТЕМА АНАЛИЗА ЭМОЦИОНАЛЬНО ОКРАШЕННОЙ РЕЧИ

На основе разработанного программно-методического комплекса была создана система анализа эмоционально окрашенной речи, которая позволяет:

классифицировать пол диктора по ЧОТ;

классифицировать возраст диктора на основе просодических параметров;

классифицировать эмоции на основе МОВ;

распознавать эмоциональную слитную речь.

Определение пола диктора осуществляется по медианному значению частоты основного тона в речевом сигнале при помощи классификатора на основе МОВ, построенного для базы из 200 мужских и 200 женских голосов. Классификация возраста диктора происходит при помощи метода k ближайших соседей [17] на основе выделяемых из речевого сигнала просодических параметров (интонированность, громкость, ритмичность, мелодичность, скорость и Джиттер), при этом каждый класс задается допустимым интервалом каждого из параметров.

В системе используется классификатор эмоций на основе МОВ и критерия Джини в качестве функции расстояния для снижения количества информативных признаков [15]. Получаемая средняя точность классификации эмоций сопоставима с лучшими известными классификаторами, а вычислительные затраты на распознавание эмоций ниже. Это позволяет использовать разработанный классификатор эмоций в составе комплексных систем анализа речевых сигналов.

Для решения задачи распознавания эмоциональной слитной речи был предложен ряд подходов повышающих точность распознавания речи в условиях слитного и эмоционального произношения. Для формирования инвариантного к эмоциям вектора признаков используется линейное предсказание и экспоненциально-логарифмическая шкала частот [18]. Также увеличить точность распознавания эмоциональной слитной речи получилось за счет использования реализованных методик декодирования спонтанной речи на основе сети спутывания и триггерной языковой модели [12], комбинированной верификации слов распознанной слитной речи на основе СММ и ДТВ [16], а также алгоритма интерактивной неконтролируемой адаптации СММ с механизмом обновления [19].

При помощи созданной системы анализа эмоционально окрашенной речи был проведен ряд экспериментов, в ходе которых были получены представленные в табл. 1 результаты.

*Таблица 1*

**Результаты тестирования системы анализа эмоционально окрашенной речи**

| Тип эксперимента | Точность, % |
|---|---|
| Классификация пола | 76,8 ± 2,9 |
| Классификация возраста | 69,4 ± 4,2 |
| Классификация эмоций | 82,7 ± 3,6 |
| Распознавание эмоциональной слитной речи | 74,6 ± 1,7 |

Ошибки классификации пола диктора вызваны сложностью вычисления частоты основного тона, которая может быть корректно определена только на вокализованных участках речи. Эффективность классификации возраста диктора сильно зависит от набора распознаваемых классов, которые задает эксперт на основе эмпирических данных. Классификация эмоций проводилась по семи состояниям (гнев, страх, отвращение, печаль, скука, радость и нейтральное эмоциональное состояние), при этом лучше распознавался гнев и печаль, а хуже всего – отвращение.

Для оценки влияния эмоциональной окраски речевого сигнала на точность распознавания необходимо модифицировать систему: на этапе параметризации речевого сигнала заменить расчет инвариантного к эмоциям вектора признаков на мел-частотные кепстральные коэффициенты, что в рамках разработанного программно-методического комплекса не составляет труда. Было установлено, что в зависимости от типа эмоции точность распознавания эмоциональной слитной речи снижается от 10 до 45 % по сравнению с таковой для нейтральной речи.

Проведенные эксперименты также показали, что точность классификации пола и возраста диктора для эмоционально окрашенной речи в среднем снижается на 15 и 35 % соответственно. Это вызвано тем, что эмоциональная речь характеризуется изменением характеристик голоса по сравнению с нейтральной речью. Таким образом, просодические параметры речи, которые рассчитываются на основе характеристик голоса, также будут изменяться для эмоционально окрашенной речи, что приводит к снижению точности классификации пола и возраста диктора.

## ЗАКЛЮЧЕНИЕ

Преимуществом использования предложенного программно-методического комплекса для разработки системы анализа эмоционально окрашенной речи является простота и минимальные временные затраты на создание конечной прикладной программы. В свою очередь экономический эффект достигается за счет возможности быстрой модификации полученной системы, чтобы она могла отвечать всем новым требованиям. Такой подход позволяет получать гибкие системы анализа речевых сигналов, что делает разработанный программно-методический комплекс особо востребованным в учебном процессе.

## БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Shenoi B. A. Introduction to digital signal processing and filter design. NJ ; Hoboken : John Wiley and Sons, Inc., 2005.

2. Hayes M. H. Statistical digital signal processing and modeling. NY ; N. Y. : John Wiley and Sons, Inc., 1996.

3. Brigham K. O. The fast Fourier transform. NJ ; Englewood Cliffs : Prentice-Hall, Inc., 1974.

4. Numerical recipes in C: the art of scientific computing / W. H. Press [et al.]. 3rd ed. N. Y. : Cambridge University Press, 2007.

5. Addison P. S. The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance. UK ; Edinburgh : Napier University, CRC Press. 2002.

6. Rabiner L., Juang B.-H. Fundamentals of speech recognition. NJ ; Upper Saddle River : Prentice-Hall, Inc., 1993.

7. Prosody [Electronic resource] : Prosody (linguistics) / Wikipedia : the free encyclopedia. Mode of access: https://en.wikipedia.org/wiki/Prosody_%28linguistics%29 (date of access: 17.08.2016).

8. Голубинский А. Н. Расчет частоты основного тона речевого сигнала на основе полигармонической математической модели // Вестн. Воронежс. ин-та МВД России. 2009. № 1. С. 48–57.

9. Nello C., Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods. UK ; Cambridge : Cambridge University Press. 2000. 204 p.

10. Myers C. S., Rabiner L. R. A comparative study of several dynamic time-warping algorithms for connected-word recognition // The Bell System Technical J. 1981. V. 60. № 7. P. 1389–1409.

11. The HTK Book / S. Young [et al.]. UK ; Cambridge University Engineering Department : Cambridge University Press. 2006.

12. Ткаченя А. В. Декодирование речи на основе триггерной сети спутывания // Электроника Инфо. 2014. № 8 (110). С. 20–23.

13. Zheng F., Zhang G., Song Z. Comparison of different implementations of MFCC // J. Computer Science and Technology. 2001. Vol. 16. № 6. P. 582–589.

14. Hermansky H. Perceptual linear predictive (PLP) analysis for speech // The J. of the Acoustical Society of America. 1990. Vol. 87, iss. 4. P. 1738–1752.

15. Классификация эмоционального состояния диктора с использованием метода опорных векторов и критерия Джини / А. В. Ткаченя [и др.] // Изв. вузов. Приборостроение. 2013. Т. 56, № 2. С. 61–66.

16. Ткаченя А. В. Верификации результатов распознавания эмоциональной слитной речи // Электроника Инфо. 2014. № 7 (109). С. 32–34.

17. Hall P., Park B. U., Samworth R. J. Choice of neighbor order in nearest-neighbor classification // The annals of statistics. 2008. Vol. 36. № 5. P. 2135–2152.

18. Ткаченя А. В. Методика формирования устойчивых к эмоциям информативных признаков для задачи распознавания речи // Изв. вузов. Приборостроение. 2015. Т. 58, № 6. С. 443–450.

19. Ткаченя А. В. Адаптация скрытых марковских моделей к распознаванию эмоционально окрашенной речи // Информатика. 2014. № 3 (43). С. 21–27.