

7. Frankenfield J. Mining Pool / J. Frankenfield. — URL: <https://www.investopedia.com/terms/m/mining-pool.asp> (date of the application: 30.05.2022). — Text : electronic.
8. Hong K. H. Bitcoin as an alternative investment vehicle / K. H. Hong. — Text : electronic // Information Technology and Management. — 2017. — № 18. — P. 265-275.
9. Kristoufek L. What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis / L. Kristoufek. — Text : electronic // PloS one. — 2015. — № 10. — P. 1-15.
10. Smith V. L. An experimental study of competitive market behavior / V. L. Smith. — Text : electronic // Journal of political economy. — 1962. — № 70. — P. 11-137.
11. Irresberger F. The public blockchain ecosystem: An empirical analysis / F. Irresberger [et al.]. — Text : electronic // NYU Stern School of Business. — 2021.

О. Ю. ПОПОВА

Тюменский государственный университет, г. Тюмень

УДК 004.93

РАЗРАБОТКА МАТЕМАТИЧЕСКОЙ МОДЕЛИ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ ЭМОЦИЙ ЧЕЛОВЕЧЕСКОЙ РЕЧИ

***Аннотация.** В публикации представлен анализ методов и практических разработок в области распознавания эмоций по аудио- и видеоданным. Предложена параметрическая математическая модель для автоматизации распознавания эмоций человеческой речи в качестве альтернативного варианта используемым искусственным нейронным сетям.*

***Ключевые слова:** распознавание эмоций, методы ИАД, математическое моделирование, распознавание образов.*

Введение. В настоящее время распознавание эмоций широко используется в различных областях: в медицине, криминалистике, разработке компьютерных игр и приложений, маркетинге, HR и образовании. Над проблемой распознавания эмоций работают ученые разных областей знаний — психологии, биологии, информатики.

В работах в области компьютерных наук по распознаванию эмоций используют набор дискретных эмоциональных состояний [1], потому что при классификации и кластеризации такой подход обеспечивает большую точность результата, чем в случаях, когда используется многовариативная классификация.

Для решения задач классификации и кластеризации при распознавании эмоций используют несколько различных математических подходов и методов, которые можно разделить на два раздела — методы интеллектуального анализа данных (ИАД) и методы математического анализа данных (в частности метод вейвлет-преобразования [2]). Больше всего исследований проведено в области использования методов ИАД. Существует несколько обзоров некоторых из существующих на данный момент программ (например, обзор в магистерской диссертации А. С. Кузнецовой [3]).

К методам ИАД относятся такие методы, как метод главных компонент, метод k -средних и искусственные нейронные сети. Сравнение эффективности этих методов для задач распознавание эмоций можно прочитать, например, в статье [4]. Минусом этих методов является отсутствие точного математического описания процесса распознавания эмоций и возможность переобучения нейросетей. Но эти методы, сейчас пробуют применяться на практике, потому что требуют минимальных исследовательских затрат после создания запрограммированной нейронной сети. К настоящему времени разработано множество компьютерных приложений, использующих технологии распознавания эмоций в той или иной мере для разных целей, в том числе предоставленные российским рынком: Affectiva, Emotient, Face++, EmoPy, EmoDetect, FindFace. Все эти приложения работают на основе искусственных нейронных сетей (ИНС), также во время проведения анализа исследований за прошедшие два года было выявлено, что отсутствуют альтернативные математические модели эмоциональных речевых сигналов. В качестве методов и в прикладной и в теоретической сфере исследований используются исключительно ИНС.

Проблема исследования. По этой причине было решено сосредоточить работу именно на разработке математической модели, позволяющей производить анализ эмоциональности речевого сигнала

без использования ИНС, чтобы облегчить и ускорить процесс вычисления.

В [5] выделены элементы эмоциональной составляющей речевого сигнала при синтезе звука. Эмоциональная составляющая речевого сигнала зависит от: темпа (модуляция длительности), тембра (модуляция формы и частотная модуляция), тона (частотная модуляция), громкости (амплитудная модуляция).

Одновременно с этим в [6, 7] указываются параметры, используемые при построении моделей распознавания эмоционального сигнала, такие как: частота основного тона (ЧОТ), форманты, среднее значение спектра, нормализованные средние значения спектра, относительно время пребывания сигнала в полосах спектра, нормализованное время пребывания сигнала в полосах спектра, вариация огибающих спектра, нормализованные величины вариации огибающих спектра, коэффициенты кроскорреляции спектральных огибающих между полосами спектра, медианное значение спектра в полосах речи, длительность фонемы, коэффициент формы сегмента, высота сегмента, амплитуда, энергия.

Именно эти параметры решено использовать для построения модели распознаваемого эмоционального речевого сигнала.

Материалы и методы. Для проверки работоспособности модели планируется использовать базу данных, в которую включены как эмоционально окрашенные, так и нейтральные записи, поэтому предполагается проверить два варианта модели: первый построен на принципе суперпозиции отклонений параметров от нейтральных значений; второй включает в себя также зависимости изменения параметров во времени. Первую модель предположительно можно будет использовать для анализа имеющихся записей. При использовании второй модели предполагается получить динамическую картину распознавания эмоций и применить при работе в «режиме реального времени» при учете того, что в начальный момент времени (t_0) будет считаться нейтральным эмоциональным состоянием.

На основе анализа работ по исследованию автоматического распознавания эмоций [8, 9] были выбраны следующие параметры для включения в параметрическую математическую модель: амплитуда

абсолютная (A), амплитуда средняя (A_{cp}), относительное изменение амплитуды (ΔA), частота основного тона (F_0), частоты первых трех формант (F_1, F_2, F_3), длительность фонемы (T_Φ), коэффициент пребывания сигнала в полосах спектра, среднее значение спектра (A_f), нормализованные средние значения спектра.

Тогда в начальном приближении математические модели будут иметь следующий вид (фрейм 1 — $[t_0, t_n]$, фрейм 2 — $[t_n, t_m]$):

$$\left\{ \begin{array}{l} |A_{max2} - A_{max1}| = D_1 \\ \frac{|\sum_{i=0}^n A_2(t_i) - \sum_{i=0}^n A_1(t_i)|}{n} = D_2 \\ \frac{|\sum_{i=0}^n (A_2(t_i) - A_{cp2}) - \sum_{i=0}^n (A_1(t_i) - A_{cp1})|}{n} = D_3 \\ |F_{02} - F_{01}| = D_4 \\ |F_{k2} - F_{k1}| = D_{4+k}, k = 1, 2, 3 \\ |A\omega_{cp2} - A\omega_{cp1}| = D_8 \\ \left| \frac{A\omega_{j2}}{A\omega_{max2}} - \frac{A\omega_{j1}}{A\omega_{max1}} \right| = D_{8+j}, j = 1, \dots, 12 \\ \left| \int_{\omega_{p-1}}^{\omega_p} (A_2(\omega) - A_1(\omega)) d\omega \right| = D_{20+p}, p = 1, \dots, 12 \end{array} \right. \quad (1)$$

где A_{max1}, A_{max2} — максимальное значение амплитуды 1 и 2 фреймов соответственно; $A_1(t_i), A_2(t_i)$ — значения амплитуды 1 и 2 фреймов соответственно в i момент времени, i принимает значения от 0 в начале фрейма и n в конечной точке фрейма, n — число тактов в фрейме; $A_{cp1} = \frac{\sum_{i=0}^n A_1(t_i)}{n}$, $A_{cp2} = \frac{\sum_{i=0}^n A_2(t_i)}{n}$ — средняя по времени амплитуда 1 и 2 фреймов соответственно; F_{01}, F_{02} — частота основного тона 1 и 2 фреймов соответственно; $F_{11}, F_{21}, F_{31}, F_{12}, F_{22}, F_{32}$ — частоты первых трех формант 1 и 2 фреймов соответственно; $A\omega_{j1}, A\omega_{j2}$ — среднее значение амплитуды j полосы спектра 1 и 2 фреймов соответственно; $A\omega_{cp1}, A\omega_{cp2}$ — среднее значение амплитуды частотного спектра 1 и 2 фреймов соответственно; $A\omega_{max1}, A\omega_{max2}$ — максимальное значение амплитуды частотного спектра 1 и 2 фреймов соответственно; ω_p — границы полос амплитудно-частотного спектра; $A_1(\omega), A_2(\omega)$ — функция зависимости амплитуды от частоты 1 и 2 фреймов соответственно; $D_i, i = 1, \dots, 30$ — критические

параметры, суперпозиция которых будет определять возможность проявления той или иной эмоции на отдельном шаге, оцениваются на основе аналогичных значений (E_i), полученных из формулы (*) для нейтрального состояния.

Относительно отклонения от известного нейтрального состояния:

$$\left\{ \begin{array}{l} |A_{max} - (A_H)_{max}| \leq C_1 \\ \frac{|\sum_{i=0}^n A(t_i) - \sum_{i=0}^n A_H(t_i)|}{n} \leq C_2 \\ \frac{|\sum_{i=0}^n (A(t_i) - A_{cp}) - \sum_{i=0}^n (A_0(t_i) - (A_H)_{cp})|}{n} \leq C_3 \\ |F_0 - (F_H)_0| \leq C_4 \\ |F_k - (F_H)_k| \leq C_{4+k}, k = 1, 2, 3 \\ |A\omega_{cp} - (A\omega_H)_{cp}| \leq C_8 \\ \left| \frac{A\omega_j}{A\omega_{max}} - \frac{(A\omega_H)_j}{(A\omega_H)_{max}} \right| \leq C_{8+j}, j = 1, \dots, 12 \\ \left| \int_{\omega_{p-1}}^{\omega_p} (A(\omega) - A_H(\omega)) d\omega \right| \leq C_{20+p}, p = 1, \dots, 12 \end{array} \right. \quad (2)$$

где A_{max} — максимальное значение амплитуды эмоционального состояния; A_{maxH} — максимальное значение амплитуды нейтрального состояния; $A(t_i)$ — значения амплитуды в i момент времени, i принимает значения от 0 в начале фрейма и n в конечной точке фрейма, n — число тактов в фрейме; $A_{cp} = \frac{\sum_{i=0}^n A(t_i)}{n}$ — средняя по времени амплитуда эмоционального состояния; $A_{cpH} = \frac{\sum_{i=0}^n A_H(t_i)}{n}$ — средняя по времени амплитуда нейтрального состояния; F_0 — частота основного тона эмоционального состояния; F_{0H} — частота основного тона нейтрального состояния; F_1, F_2, F_3 — частоты первых трех формант эмоционального состояния; F_{1H}, F_{2H}, F_{3H} — частоты первых трех формант нейтрального состояния; $A\omega_j$ — среднее значение амплитуды j полосы спектра; $A\omega_{cp}$ — среднее значение амплитуды частотного спектра эмоционального состояния; $A\omega_{cpH}$ — среднее значение амплитуды частотного спектра нейтрального состояния; $A\omega_{max}$ — максимальное значение амплитуды частотного спектра

эмоционального состояния; $A\omega_{\text{нmax}}$ — максимальное значение амплитуды частотного спектра нейтрального состояния; $A\omega_p$ — границы полос амплитудно-частотного спектра; $A(\omega)$ — функция зависимости амплитуды эмоционального состояния от частоты; $A_n(\omega)$ — функция зависимости амплитуды эмоционального состояния от частоты; $C_i, i = 1, \dots, 30$ — критические параметры, суперпозиция которых будет определять возможность проявления той или иной эмоции на отдельном шаге, рассчитываются на основе разницы значений D_i и E_i из формулы (1).

В ходе эксперимента планируется установить параметры D_i, C_i , что позволит в дальнейшем скорректировать модель при учете параметров для отдельных эмоций и оформить отдельные модели для различных эмоциональных состояний с учетом разницы суперпозиций критических параметров. После этого возможна проверка уточненной модели на другой базе данных и дальнейшая доработка при работе с необработанными данными.

Для проверки и корректировки предложенной математической модели было решено использовать Аудиовизуальную базу данных эмоционально записанных речи и песен Райерсона (RAVDESS) [10].

Изначально были взяты только файлы с записями без эмоций и две противоположные эмоции — радость и грусть для оптимизации задачи распознавания. Было решено для экспериментальной выборки взять 71% записей. Для нейтрального состояния — 68 записей. Для радости/грусти — по 136 записей. Оставшиеся записи должны будут использоваться для проверки модели.

Подготовка каждой записи включала удаление момента тишины в начале и в конце записи. Оставшаяся часть была поделена на 19 равных фреймов (с перекрытием 50%) примерно по 9600 отчетов.

Результаты. Обработка речевых записей начата с нейтральной записи первого актера, чтобы иметь возможность разобраться с механизмом обработки сигнала и получением необходимых параметров.

Сигнал был очищен от начального и конечного участков тишины — вид получившегося сигнала представлен на рис. 1.

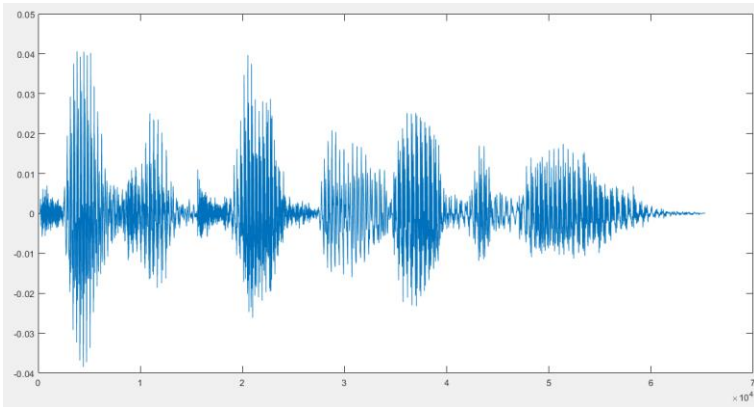


Рис. 1. Запись 03-01-01-01-01-01-01 без моментов тишины

Вторым шагом полученная запись была разделена на 19 фреймов (с областью перекрытия 50%). Вид одного из фреймов представлен на рис. 2.

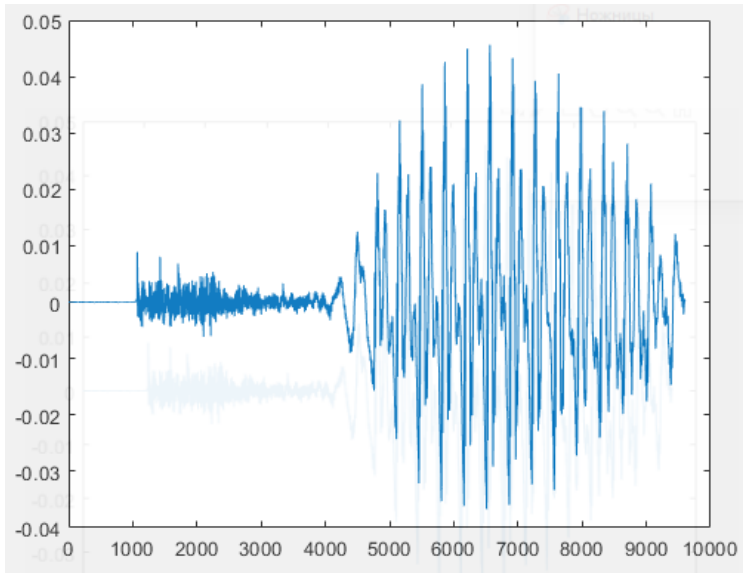


Рис. 2. График отдельного фрейма

По данным этих фреймов будут получены значения параметров: A_{max} — максимальное значение амплитуды фрейма; A_{cp} — среднее арифметическое значение амплитуды фрейма; ΔA_{cp} — среднее математическое отклонение значений амплитуды фрейма.

Третьим шагом были получены амплитудно-частотные спектры каждого фрейма. АЧХ того же фрейма приведена на рис. 3.

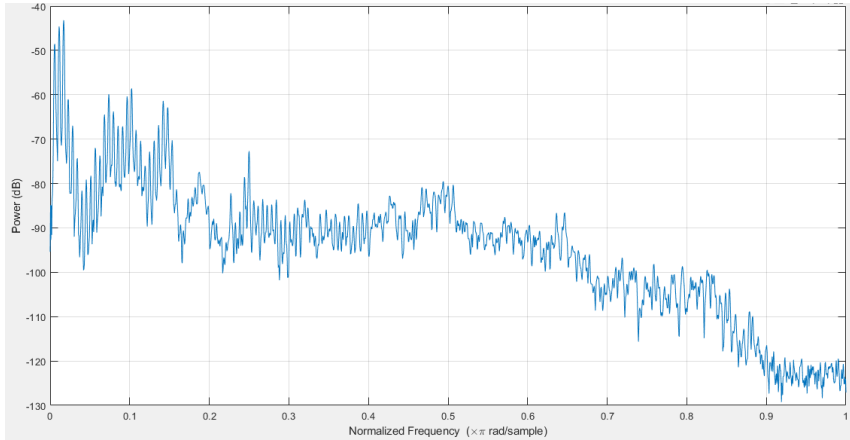


Рис. 3. АЧХ одного из фреймов

На основе данных этих спектров получены значения следующих параметров (в зависимости от номера фрейма): ЧОТ — частота основного тона, первый глобальный максимум спектра; F_1 , F_2 , F_3 — первая, вторая и третья форманты, соответственно первый, второй и третий максимумы АЧХ; $A_{\omega_{max}}$ — максимальное значение амплитуды АЧХ; $A_{\omega_{cp}}$ — среднее значение амплитуды АЧХ; $(A_{\omega_{n1}}, A_{\omega_{n12}})$ — среднее значение амплитуды в двенадцати полосах спектра; $(A_{\omega_{инт1}}, A_{\omega_{инт12}})$ — значения интегралов в полосах спектра.

Пример зависимости величины формант от номера фрейма представлен на рис. 4.

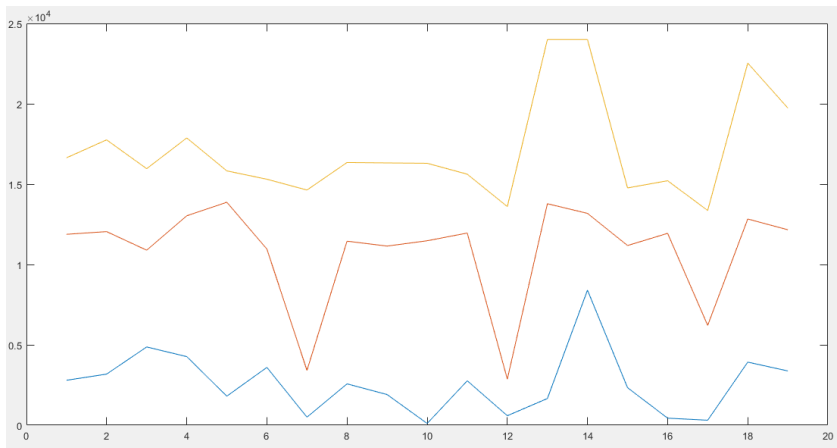


Рис. 4. Зависимость величины форманты от номера фрейма (желтый — первая форманта, оранжевый — вторая форманта, синий — третья форманта)

Заключение. Была разработана основа для параметрической модели распознавания эмоций. Она включает в себя две составляющие — первая оценивает изменение речевого сигнала, вторая сравнивает это изменение с эмоциональным проявлением.

Начата работа по численному выявлению предложенных параметров на основе анализа базы данных RAVDESS.

Намечена дальнейшая научная деятельность по теме исследования: обработать речевые записи базы данных, скорректировать предложенную математическую модель, проверить работоспособность модели на оставшейся выборке, проверить работоспособность модели на записи речевого сигнала, не принадлежащего базе данных.

СПИСОК ЛИТЕРАТУРЫ

1. Давыдов А. Г. Классификация эмоционального состояния диктора по голосу: проблемы и решения / А. Г. Давыдов, В. В. Киселев, Д. С. Кочетков. — Текст : непосредственный // Труды международной конференции «Диалог 2011». — Москва : РГТУ, 2011. — С. 178-185.

2. Бредихин А. И. Распознавание эмоций в человеческой речи с использованием вейвлетов / А. И. Бредихин. — Текст : электронный // Научное сообщество студентов XXI столетия. Технические науки : сборник статей по материалам LXII международной студенческой научно-практической конференции. — 2018. — № 2 (61). — URL: [https://sibac.info/archive/technic/2\(61\).pdf](https://sibac.info/archive/technic/2(61).pdf) (дата обращения: 10.05.2022).
3. Кузнецова А. С. Нейросетевые алгоритмы распознавания эмоционального состояния человека по изображениям его лица : дипломная работа (диссертация магистра информатики и вычислительной техники) / А. С. Кузнецова ; Томский политехнический университет. — Томск, 2019. — 108 с. — Текст : непосредственный.
4. Полякова А. С. Комбинирование подходов кластеризации и классификации для задачи распознавания эмоций по речи / А. С. Полякова, М. Ю. Сидоров, Е. С. Семенкин. — Текст : непосредственный // Вестник СибГАУ. — 2016. — Т. 17, № 2. — С. 335-342.
5. Лобанов Б. М. Компьютерный синтез и клонирование речи / Б. М. Лобанов, Л. И. Цирульник. — Минск : Белорусская наука, 2008. — 316 с. — Текст : непосредственный.
6. Розалиев В. Л. Моделирование эмоциональных реакций пользователя при речевом взаимодействии с автоматизированной системой : 05.13.01 : дис. ... канд. техн. наук / В. Л. Розалиев ; ВГТУ. — Волгоград, 2009. — 163 с. — Текст : непосредственный.
7. Сидоров К. В. Анализ признаков эмоционально-окрашенной речи / К. В. Сидоров, Н. Н. Филатова. — Текст : непосредственный // Вестник Тверского государственного технического университета. — 2012. — Вып. 20. — С. 26–31.
8. Томчук К. К. Сегментация речевых сигналов для задач автоматической обработки речи : дис. ... канд. техн. наук / К. К. Томчук ; Санкт-Петербургский государственный университет аэрокосмического приборостроения. — Санкт-Петербург, 2017. — 197 с. — Текст : непосредственный.
9. Дворянкин С. В. Применение фазовых характеристик голосовых вокализмов в решении задач защиты речевой информации / С. В. Дворянкин, А. О. Антипенко. — Текст : электронный // Безопасность информационных технологий. — 2021. — Vol. 28, № 2. — P. 21-33. — DOI : <http://dx.doi.org/10.26583/bit.2021.2.02> (дата обращения: 10.05.2022).

10. Livingstone S. R. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English / S. R. Livingstone, F. A. Russo. — 2018. — PLoS ONE 13(5): e0196391. — DOI: <https://doi.org/10.1371/journal.pone.0196391> (date of the application: 10.05.2020). — Text : electronic.

К. С. РЕЗАНОВ, Н. К. ПРОТОЗАНОВ

*Российский государственный университет нефти и газа (НИУ)
имени И. М. Губкина, г. Москва*

УДК 621.643.053

ОЦЕНКА ЭФФЕКТИВНОСТИ ПРИМЕНЕНИЯ ДОЖИМНЫХ КОМПРЕССОРОВ ПРИ ПОЛУЧЕНИИ СПГ НА ГРС

Аннотация. При получении малотоннажного СПГ на ГРС существует ряд проблем, осложняющих работу установок получения, например сезонный перепад давления на входе в установку, что может быть решено путем установки дополнительного компрессора. В работе представлены модели расчета в программном комплексе “Unisim Desigh”, на основании которых в программе проведен расчет для различных циклов получения СПГ на ГРС.

Ключевые слова: СПГ, ГРС, компрессор, компьютерное моделирование.

Введение. На данный момент одним из самых перспективных методов развития малотоннажного СПГ является использование газораспределительных станций. В РФ имеется опыт производства и применения СПГ на основе использования отечественных технических средств и технологических процессов, а также зарубежных достижений [1].

Использование ГРС для сжижения природного газа эффективно в первую очередь за счет сокращения оборудования по причине использования перепада давления на входе и выходе из станции. Именно эту технологию целесообразно принять в качестве основной при организации производственной инфраструктуры СПГ комплексов. Процесс сжижения газа на ГРС происходит при редуцировании