

**АНАЛИЗ СОСТОЯНИЯ ЧЛЕНА ЭКИПАЖА ВОЗДУШНОГО СУДНА
ПО ЕГО РЕЧИ НА БАЗЕ ГАУССОВЫХ МОДЕЛЕЙ СМЕСЕЙ**© 2021 Н.А. Андриянов¹, В.Е. Дементьев²¹ Финансовый университет при Правительстве Российской Федерации, г. Москва, Россия² Ульяновский государственный технический университет, г. Ульяновск, Россия

Статья поступила в редакцию 10.12.2020

Работа посвящена исследованию эффективности применения моделей гауссовых смесей для распознавания аномальных отклонений в речи диктора. Предложено практическое применение разрабатываемых алгоритмов для выявления эмоционального состояния члена экипажа по произнесенной им фразе. В качестве основного критерия для различения с помощью модели гауссовых смесей используются спектральные характеристики речевого сигнала. В связи с достаточно небольшим шагом дискретизации по частоте и соответственно с наличием 255 частотных составляющих в спектре сигнала предложено сжатие спектра до 10 составляющих. Такой подход позволил сократить число ключевых параметров в гауссовой модели до 10, что, в свою очередь, позволило упростить процесс анализа при построении многомерных распределений. Для оценки качества предлагаемого алгоритма были записаны тестовые фразы. При этом имитировались различные психологические состояния диктора. Использовались как простые нерегламентированные речевые конструкции, так и сообщения, регулируемые в порядке Федеральных авиационных правил при проведении радиообмена в гражданской авиации на территории Российской Федерации. С учетом ограничений на предварительные знания модели и кластеризации по спектральным характеристикам все записи модели были выполнены одним диктором. Были рассмотрены три класса эмоционального состояния диктора. На выходе система распознавания ставила такие метки, как спокойное состояние, усталое состояние, стрессовое состояние. Различные состояния искусственно имитировались в процессе подготовки данных. На тестовой выборке из 48 сообщений гауссова модель из 3 компонент и 10 параметров без предварительного обучения сразу позволила достичь результата порядка 65%, в то время как вероятность распознать верный класс при 3-х равных классах априори составляет 33%. В качестве дальнейших исследований предложено применение предварительного обучения с использованием нейронных сетей или корреляционных алгоритмов. Такой подход позволит выполнять дальнейшую кластеризацию на более глубоком уровне, когда, например, определяется пол диктора, определяется типовое сообщение радиообмена, а затем уже выявляется эмоциональное состояние диктора.

Ключевые слова: спектральный анализ, распознавание эмоционального состояния речи, интеллектуальный анализ данных, обеспечение безопасности полетов, гауссова модель смесей

DOI: 10.37313/1990-5378-2021-23-1-97-102

*Работа выполнена при поддержке гранта РФФИ, Проект №19-47-730011.***ВВЕДЕНИЕ**

Обеспечение безопасности полетов является главным приоритетом современной гражданской авиации. Сложно переоценить роль этой задачи в условиях, когда на борту самолета находится множество людей. Однако по статистике [1] около 80% авиакатастроф в Российской Федерации происходят по причине человеческого фактора, и лишь 20% случаев приходится на отказы авиационной техники. Понятно, что помимо глобальных факторов, связанных с непра-

вильным поведением экипажа в критических ситуациях, нехваткой летной практики, важную роль играет и психологический фактор. В каком состоянии пилот начинает рейс? Потенциально выявление и предупреждение некоторых отклонений в состоянии пилота в ходе предполетной проверки может дополнительно снизить риски. В мире существуют специальные подходы и организации, способствующие реализации системы мониторинга состояния экипажа [2-4]. Однако в настоящее время не существует достаточно эффективной системы выявления усталости или состояний психологического отклонения у членов экипажа. Подобные системы, например, реализуются при мониторинге состояния водителя с помощью искусственного интеллекта и задач компьютерного зрения [5-7]. В частности, решаются задачи распознавания закрытых

Андриянов Никита Андреевич, кандидат технических наук, доцент Департамента анализа данных и машинного обучения. E-mail: nikita-and-nov@mail.ru

Дементьев Виталий Евгеньевич, доктор технических наук, доцент, заведующий кафедрой «Телекоммуникации». E-mail: dve@ulntc.ru

глаз, сигнализирующих об усталости водителя. Интеллектуальный анализ данных также может быть успешно применен и для распознавания отклонений в голосе членов экипажа воздушного судна (ВС). Попытка имитации диалогов пилота в различных условиях рассмотрена в работе [8]. Однако авторами рассматривается лишь моделирование ограниченного числа фраз сообщений фразеологического радиообмена. При этом имитируется голос уже непосредственно с борта ВС. Более того, получение достаточной обучающей выборки в данной ситуации довольно критично, поскольку нужно записывать голос пилота и отдавать на анализ экспертам. Адекватная оценка классификации, проведенной экспертами, будет возможна лишь спустя время после выполнения полета. При этом будет требоваться достаточно сложный дополнительный анализ. Заметим, что одним из вариантов быстрой кластеризации данных является применение моделей гауссовых смесей [9-10]. При этом не требуется предварительно знать, к какому классу относится то или иное сообщение. Понятно, что такая система будет уступать хорошо обученным под выявление важных параметров отклонения психологического состояния, но может значительно сократить время реализации в условиях подготовки достаточной обучающей выборки.

Целью данной работы является исследование применения гауссовых моделей смесей для выявления стандартных и аномальных сообщений одного и того же диктора по спектральному представлению данного сигнала.

Таким образом, важно, что при подходе анализа спектральных характеристик речевые сообщения при кластеризации без учителя относятся к одному и тому же диктору. Действительно, спектр сообщений, произнесенных различными дикторами, особенно разного пола, будет сильно отличаться, и при автоматическом выборе параметров для классификации наиболее вероятным будет разделение именно по голосу.

МЕТОДЫ ИССЛЕДОВАНИЙ

Задача исследования заключалась в оценке состояния человека по голосу. Для этого необходимо было подготовить исходную базу данных для анализа. Одним и тем же диктором произносились, как фразы по установленной форме радиообмена, так и свободные вольные фразы. Дополнительный шум искусственно при записи не добавлялся. Дополнительной предобработки с записями по эквалайзингу и фильтрации также не выполнялось. Для удобства оценки в отсутствие экспертов в области психологии все выражения записывались трижды в различных состояниях. При этом предпринималась попытка

повторения одной и той же фразы в стандартном режиме, в заторможенном режиме с паузами (имитация усталости), в обрывистом, беспокойном режиме (имитация психологического отклонения). Таким образом, было создано 3 класса сообщений: стандарт, усталость, нервозность.

Использовалась стандартная бесплатная программа для работы с аудио сигналами «AudaCity». Данное программное обеспечение позволяет формировать спектр сигнала в диапазоне от 86 Гц до 22 кГц. Однако данный подход обеспечивает 255 составляющих частотного спектра, что достаточно много для модели гауссовых смесей. Простым усреднением было выполнено схлопывание спектра до 10 компонент. При этом частота выбиралась в середине участка, а значение на заданной частоте получалось по выражению (1)

$$\frac{1}{N} \sum_{i=j}^{j+N-1} S_i, \quad (1)$$

где N – число компонент для усреднения (25 для первых 9 новых компонент, 30 для последней новой компоненты), j – номер частотной компоненты, с которой начинается усреднение, S_j – уровень сигнала на j -й частотной составляющей.

Таким образом, после усреднения получается 10 компонент. Значения данных компонент и будут являться параметрами модели (в 10 измерениях). На основе 10-мерной плотности распределения вероятностей будут формироваться определенные классы. При этом будем считать, что заранее известно точное число классов, а именно 3.

Рассмотрим на примере одномерного случая модель гауссовых смесей более подробно. Поскольку случайные значения сигналов на частотных составляющих наиболее точно могут быть описаны с помощью нормального распределения, применение такой модели целесообразно для решения поставленной задачи. Пусть классифицируемый объект описывается одним параметром x_0 , например, уровнем сигнала на частоте F_0 . При аппроксимации случайной величины x_0 с помощью гауссова распределения можно перейти к выражению плотности распределения вероятностей (ПРВ)

$$f(x_0) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_0 - m)^2}{2\sigma^2}\right), \quad (2)$$

где m – математическое ожидание или среднее значение случайной величины x_0 , а σ – среднеквадратичное отклонение случайной величины x_0 .

Теперь допустим, что, кроме значения сигнала на частотной составляющей F_0 , добавляются

еще значения сигнала на частотных составляющих F_1, F_2, \dots, F_{k-1} , и эти значения описываются случайными величинами x_1, x_2, \dots, x_{k-1} . Тогда размерность ПРВ увеличивается до k . Однако основным преимуществом нормального распределения (2) является удобство его представления в многомерной форме (3)

$$f(\bar{x}) = \frac{1}{(2\pi)^{n/2} \sqrt{\det V_x}} \exp\left(-\frac{1}{2}(\bar{x} - \bar{m})^T V_x^{-1} (\bar{x} - \bar{m})\right), \quad (3)$$

где $\bar{x} = (x_0, x_1, \dots, x_{k-1})^T$, $\bar{m} = (m_0, m_1, \dots, m_{k-1})^T$, $V_x = M\{(\bar{X} - \bar{m})(\bar{X} - \bar{m})^T\}$, V_x^{-1} – матрица, обратная к V_x .

Поскольку все параметры учитываются лишь в степени экспоненты, то выражение (3) может быть заменено произведением выражений для одномерных ПРВ (2) в том случае, если между случайными величинами отсутствует корреляция. Следует отметить, что для рассматриваемой задачи такое упрощение недопустимо, поскольку существует связь между составляющими частотного спектра, близко расположенными к друг другу.

На рис. 1 представлен пример плотности распределения вероятностей для нормального

распределения при нулевом математическом ожидании и единичной дисперсии.

Как видно из рис. 1, величине x соответствует наиболее вероятное значение, равное 0. Поскольку представлен одномерный вариант, то описывается лишь точка на оси OX . При добавлении второго параметра каждый объект будет уже описываться точкой на плоскости, а ПРВ будет представляться в трехмерном пространстве. Таким образом, можно использовать сечения для анализа 10 параметров одного и того же объекта (фразы).

Модель гауссовых смесей применяется для кластеризации, когда разным значениям случайной величины x соответствуют разные классы объектов. На рис. 2 показана плотность распределения такой модели, содержащей 3 типа объектов. Пунктиром показаны отдельные распределения для таких объектов, характеризующихся одним параметром.

При этом суммарная фигура строится таким образом, чтобы ее площадь, как и для отдельных ПРВ компонент, была равна 1. Наличие нескольких пиков на графике ПРВ позволяет выполнять автоматическую кластеризацию путем простого расчета расстояния

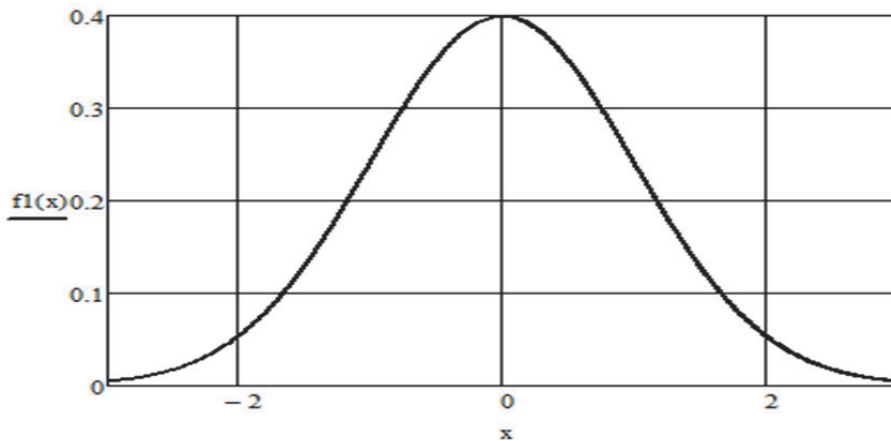


Рис. 1. Одномерное гауссово распределение

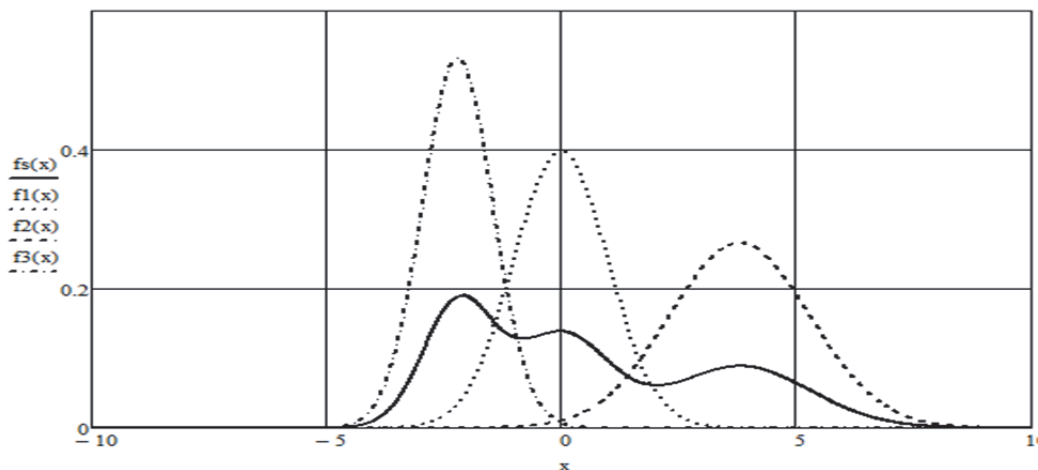
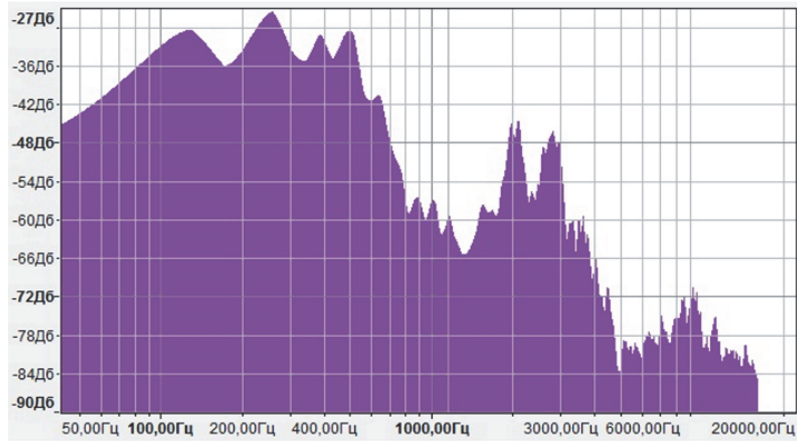


Рис. 2. Модель гауссовых смесей из трех компонент

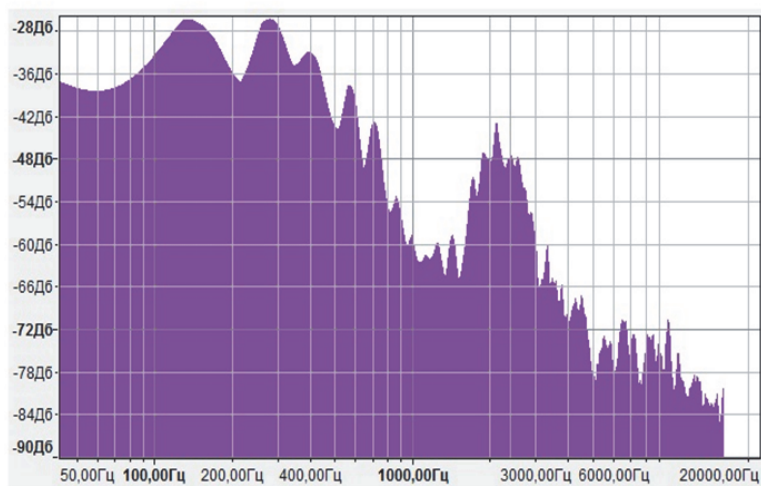
точки от каждого пика и выборе минимального расстояния для определения класса, т.е., к какому пику ближе значение параметра неизвестного объекта, к тому классу его и будет относить модель смесей.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ И ИХ ОБСУЖДЕНИЕ

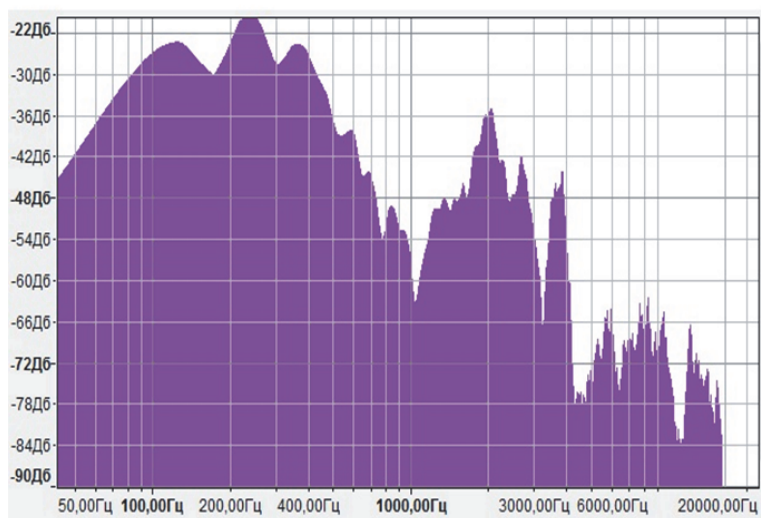
Для получения анализируемых данных использовалась стандартное звукозаписывающее



а



б



в

Рис. 3. Спектры исследуемых фраз:
а – спокойное произнесение, б – усталое произнесение, в – стрессовое произнесение

устройство типа «TASCAM время DR-40». Фразы произносились одним и тем же диктором в обычном режиме, а также в режиме стрессового состояния (после комплекса упражнений) и в режиме усталости (перед сном в конце дня). Полученные в ходе экспериментов записи были проанализированы в «AudaCity» для спектрального представления, затем спектр был разделен на 10 участков, для которых происходило усреднение.

Фразы использовались для перспективного выявления усталости у экипажа. И стандартные записи могут быть использованы как эталоны при дальнейшем анализе на борту, выявляющем расхождения от эталонов. Выбранные для эксперимента речевые сообщения представляют как стандартные фразы радиообмена, так и фразы более свободной формы.

Далее приведены 8 фраз, используемые при анализе.

- 1) MAY DAY, MAY DAY, MAY DAY (Фраза №1);
- 2) Внимание экипаж, сдвиг ветра, уходим на второй круг (Фраза №2);
- 3) Внимание экипаж, карта аварийных проверок, пожар правого двигателя (Фраза №3);
- 4) 1013 ГПа установлено, набираю эшелон 80 (Фраза №4);
- 5) Здравствуйте, как дела? (Фраза №5);
- 6) Добрый день, температура в пункте прибытия плюс двадцать (Фраза №6);
- 7) Я командир воздушного судна Diamond 40NG (Фраза №7);
- 8) Предполетная проверка связи. Раз, два, три. (Фраза №8).

Каждая фраза была записана по 2 раза в каждом состоянии. Это позволило получить выборку из 48 звуковых сообщений. Соответственно в исходной выборке было по 16 сообщений каждого класса.

На рис. 3 представлены примеры спектров, полученных при произнесении Фразы №1 в различных режимах.

Из рис. 3 видно, что различные условия произнесения фраз приводят к появлению различных спектров. Это и было использовано при применении гауссовых моделей смесей.

В табл. 1 показаны результаты распознавания для всех 3-х классов.

Анализ полученных результатов показывает, что применение гауссовых моделей смесей для данных фраз обеспечило долю правильных распознаваний среди 3 классов порядка 65%.

ЗАКЛЮЧЕНИЕ

Таким образом, в работе исследованы алгоритмы автоматической кластеризации состояния диктора по произнесенному сообщению на базе моделей гауссовых смесей. Для тестовой выборки был получен средний результат 65%. Добиться повышения качества можно, дополнительно используя корреляционный анализ для предварительного определения того, какая фраза произнесена. А затем из отсортированных фраз уже производить более глубокую кластеризацию. Это и составляет планы будущих исследований.

СПИСОК ЛИТЕРАТУРЫ

1. Авиакатастрофы и человеческий фактор. Инфографика. Электронный ресурс. Режим доступа: https://jets.ru/business-aviation/aviakatastrofy_i_chelovecheskiy_faktor_infografika_ (дата обращения: 29.11.2020).
2. *Stephens Ch., Harrivel A., Prinzel L., Comstock R., Abraham N., Pope A., Wilkerson J., Kiggins D.* Crew State Monitoring and Line-Oriented Flight Training for Attention Management // International Symposium on Aviation Psychology, 2017, p. 1-6
3. *Harrivel, A. R., Stephens, C. L., Milletich, R. L., Heinich, C. M., Last, M. C., Napoli, N. J., Pope, A. T.* (2017). Prediction of Cognitive States during Flight Simulation using Multimodal Psychophysiological Sensing. Paper presented at the American Institute of Aeronautics and Astronautics, SciTech, Grapevine, Texas.
4. Электронный ресурс. Commercial Aviation Safety Team. (2014a). Airplane State Awareness Joint Safety Analysis Team Interim Report. Режим доступа: http://www.skybrary.aero/index.php/Commercial_Aviation_Safety_Team_%28CAST%29_Reports (дата обращения: 29.11.2020).
5. Электронный ресурс. Режим доступа: <https://www.sowa.pro/> (дата обращения: 29.11.2020).

Таблица 1. Процент правильных распознаваний

№ п/п	Класс	Спокойный	Усталый	Стрессовый	Доля
1	Спокойный	10	3	3	62,5%
2	Усталый	3	11	2	68,75%
3	Стрессовый	2	4	10	62,5%
	Общая:				64,6%

6. *Гладких А.А., Андриянов Н.А., Волков А.К.* Сравнительный анализ результатов трансфера обучения нейросетей при решении задачи распознавания объектов на рентгеновских снимках багажа // Современные проблемы проектирования, производства и эксплуатации радиотехнических систем. Сборник научных трудов. Ульяновск, 2020. С. 107-110.
7. *Andriyanov N.A., Volkov A.K., Volkov A.K., Gladkikh A.A., Danilov S.D.* Automatic X-ray image analysis for aviation security with limited computing resources // IOP Conference Series: Materials Science and Engineering. Krasnoyarsk Science and Technology City Hall of the Russian Union of Scientific and Engineering Associations. 2020. С. 52009.
8. *Андриянов Н.А., Исаев М.В.* Спектральный анализ сообщений экипажа воздушного судна, сделанных в различных условиях // Современные проблемы проектирования, производства и эксплуатации радиотехнических систем. Сборник научных трудов. 2019. С. 118-123.
9. *Андриянов Н.А., Дементьев В.Е.* Обзор применения гауссовых моделей в статистической обработке данных // Современные проблемы проектирования, производства и эксплуатации радиотехнических систем. Сборник научных трудов. Ульяновск, 2020. С. 91-96.
10. *Рахманенко И.А.* Алгоритмы и программные средства верификации диктора по произвольной фразе: Диссертация на соискание ученой степени кандидата технических наук, Томск, ТУСУР, 2017. - 111 с.

ANALYSIS OF THE STATE OF AIRCRAFT CREW MEMBER BY THE SPEECH USING GAUSSIAN MODELS OF MIXTURES

© 2021 N.A. Andriyanov¹, V.E. Dementiev²

¹Financial University under the Government of the Russian Federation, Moscow, Russia

²Ulyanovsk State Technical University, Ulyanovsk, Russia

The work is devoted to the study of the effectiveness of the application of models of Gaussian mixtures for the recognition of abnormal deviations in the speaker's speech. The practical application of the developed algorithms for revealing the emotional state of the crew member by the phrase uttered by such crew member is proposed. The spectral characteristics of the speech signal are used as the main criterion for distinguishing using the Gaussian mixture model. In connection with a rather small sampling step in frequency and, accordingly, with the presence of 255 frequency components in the signal spectrum, it is proposed to compress the spectrum to 10 components. This approach made it possible to reduce the number of key parameters in the Gaussian model to 10, which, in turn, made it possible to simplify the analysis process when constructing multivariate distributions. To assess the quality of the proposed algorithm, test phrases were recorded. At the same time, various psychological states of the speaker were imitated. We used both simple unregulated speech structures and messages regulated in accordance with the Federal Aviation Rules when conducting radio exchange in civil aviation on the territory of the Russian Federation. Taking into account the limitations on the prior knowledge of the model and clustering by spectral characteristics, all recordings of the model were made by one speaker. Three classes of the speaker's emotional state were considered. At the output, the recognition system put such marks as a calm state, a tired state, a stressful state. Various states were artificially simulated during data preparation. On a test sample of 48 messages, a Gaussian model of 3 components and 10 parameters without preliminary training immediately allowed to achieve a result of about 65%, while the probability of recognizing the correct class with 3 equal classes a priori is 33%. As further research, it is proposed to apply preliminary training using neural networks or correlation algorithms. This approach will allow further clustering at a deeper level, when, for example, the gender of the speaker is determined, a typical message of the radio exchange is determined, and then the emotional state of the speaker is revealed. *Keywords:* spectral analysis, speech emotional state recognition, data mining, flight safety, Gaussian mixture model.

DOI: 10.37313/1990-5378-2021-23-1-97-102

Nikita Andriyanov, Candidate of Technics, Assistant Professor at the Data Analysis and Machine Learning Department. E-mail: nikita-and-nov@mail.ru

Vitaly Dementiev, Doctor of Technics, Head of Telecommunication Department. E-mail: dve@ulntc.ru