

DOI: 10.32347/2412-9933.2020.41.95-100

УДК 004.056.5

Терейковская Людмила Алексеевна

Кандидат технических наук, доцент, доцент кафедры информационных технологий проектирования и прикладной математики, orcid.org/0000-0002-8830-0790

Киевский национальный университет строительства и архитектуры, Киев

НЕЙРОСЕТЕВАЯ МОДЕЛЬ РАСПОЗНАВАНИЯ ЭМОЦИЙ ПО ГОЛОСУ

Аннотация Разработаны средства распознавания эмоционального состояния диктора. Показана перспективность использования нейронных сетей для анализа фиксированных фрагментов голосового сигнала. Установлена необходимость адаптации вида и параметров нейросетевой модели к условиям задачи распознавания эмоций по голосу. В результате проведенных исследований определено, что в задаче распознавания эмоций диктора по голосовым фрагментам фиксированной продолжительности целесообразно использовать двухслойный перцептрон, входные параметры которого ассоциируются с мел-кепстральными коэффициентами, характеризующими каждый из квазистационарных фрагментов анализируемого голосового сигнала, а выходные параметры соответствуют распознаваемым эмоциям диктора. Целесообразность использования двухслойного перцептрона подтверждена с помощью компьютерных экспериментов. Определено, что направления дальнейших исследований связаны с определением количества мел-кепстральных коэффициентов, которое является достаточным для описания одного квазистационарного фрагмента, и адаптацией параметров двухслойного перцептрона к условиям распознавания при воздействии помех различного вида.

Ключевые слова: эмоции; распознавание эмоций; голосовой сигнал; нейросетевая модель; мел-кепстральные коэффициенты; квазистационарный фрагмент

Введение

Мировой опыт свидетельствует о том, что одной из наиболее важных тенденций развития информационных систем общего и специального назначения является внедрение в них средств распознавания эмоционального состояния. Так, известны удачные попытки применения подобных средств в системах распознавания воздействия на подсознание в мультимедийных сообщениях средств массовой информации при ведении информационных войн, для диагностирования психических заболеваний в целях определения впечатлений клиентов от новой продукции, в индустрии развлечений и в системах управления умным домом. Кроме того, вызывает интерес применение средств распознавания эмоций в системах дистанционного образования для оптимизации учебной нагрузки. Именно распознавание эмоций заложено в основе функционирования большинства известных систем распознавания лжи (полиграфов). Таким образом, внедрение средств распознавания эмоций в интерфейс информационных систем общего назначения является одним из современных направлений их совершенствования.

В настоящее время наиболее распространенными являются средства распознавания эмоций на основе анализа геометрии лица человека. Однако

существенным препятствием использования таких средств является необходимость получения качественного изображения лица человека. Поэтому одним из основных направлений развития средств распознавания эмоций является применение в них анализаторов других биометрических характеристик, одной из которых является голос человека [12; 16]. Заметим, что перспективность использования голосовых анализаторов объясняется тем, что сегодня средства регистрации голоса являются неотъемлемой частью большинства известных информационных систем. Вместе с тем, практический опыт и результаты научно-прикладных работ [1; 10; 14] указывают на необходимость существенной модернизации таких анализаторов в направлении уменьшения ресурсоемкости, увеличения точности распознавания, расширения номенклатуры распознаваемых эмоций, сокращения срока разработки и повышения уровня адаптации к другим условиям применения, что и обуславливает актуальность исследований в этом направлении.

Анализ современных исследований и публикаций

Результаты анализа научно-практических работ свидетельствуют о том, что в настоящее время в основе распространенных систем распознавания эмоций диктора положены математические модели,

базирующиеся на байесовском подходе [2], скрытых марковских процессах [1], Support Vector Machines [6; 7], а также теории нейронных сетей [3; 5; 13]. При этом на вход системы распознавания подаются параметры, характеризующие некоторый участок голосового сигнала, регистрируемый с частотой дискретизации от 8 до 96 кГц. В работах [4; 11; 18] показана возможность эффективного распознавания личности диктора на фиксированных участках голосового сигнала продолжительностью 5 – 15 с. Сложившийся подход к процедуре обработки голосового сигнала состоит в использовании кратковременного анализа. То есть голосовой сигнал разбивается на временные окна фиксированного размера (квазистационарные фрагменты), на которых параметры сигнала не меняются. Обычно размер квазистационарного фрагмента выбирается в пределах 10 – 30 мс [11]. Для более точного представления сигнала между фрагментами делают перекрытие, равное половине длины окна. Затем к каждому фрагменту применяются алгоритмы извлечения признаков, в качестве которых, как правило, используются так называемые мел-кепстральные коэффициенты (MFCC) [6; 9]. Расчет MFCC заключается в последовательной реализации следующих шагов:

– входной дискретный голосовой сигнал фильтруется с помощью выражений:

$$z(n) = x(n) - 0,9x(n-1); \quad (1)$$

$$x(n) = z(n) \left[0,54 - 0,46 \cos \left((i-6) \frac{2\pi}{180} \right) \right], \quad (2)$$

где $x(n)$ – дискретное значение амплитуды голосового сигнала при n -й регистрации;

– рассчитывается спектр голосового сигнала.

Для этого в большинстве рассмотренных работ используется метод дискретного преобразования Фурье:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-i2\pi kn/N}, \quad (3)$$

где N – количество регистраций голосового сигнала.

Для сглаживания спектрограммы на границах квазистационарных фрагментов к значениям, полученным с помощью выражения (4), применяется оконная функция Хемминга:

$$X(k) = X(k)H(k), \quad (4)$$

где $H(k) = [0,54 - 0,46 \cos(2\pi k/N - 1)]$ – функция Хемминга.

– используя выражения (5) – (8), рассчитывают параметры фильтров:

$$B(b) = 1125 \ln(1 + b/700); \quad (5)$$

$$B^{-1}(b) = 700(e^{b/1125} - 1); \quad (6)$$

$$f(m) = \left(\frac{N}{F} \right) B^{-1} \left(B(f_1) + m \frac{B(f_h) - B(f_1)}{M+1} \right); \quad (7)$$

$$H_m(k) = \begin{cases} 0, & \text{if } k < f(m-1) \\ \frac{k - f(m-1)}{f(m) - f(m-1)}, & \text{if } f(m-1) \leq k < f(m) \\ \frac{f(m+1) - k}{f(m+1) - f(m)}, & \text{if } f(m) \leq k \leq f(m+1) \\ 0, & \text{if } k > f(m+1) \end{cases}, \quad (8)$$

где m – номер фильтра; M – количество фильтров (MFCC); F – частота дискретизации.

Используя выражения (9), (10), рассчитывают значения MFCC:

$$S(m) = \ln \left(\sum_{k=0}^{N-1} X(k)^2 H_m(k) \right), 0 \leq m \leq M; \quad (9)$$

$$C(n) = \sum_{m=0}^{M-1} S(m) \cos(\pi n(m+0,5)), 0 \leq n \leq M. \quad (10)$$

Также анализ литературных источников позволяет утверждать, что основные усилия в области создания малоресурсных клиентских средств распознавания эмоций диктора связаны с разработкой нейросетевых моделей. При этом в проанализированной литературе нет единого мнения о том, какой вид нейросетевой модели целесообразно использовать для распознавания эмоций диктора. Также недостаточно полно освещен вопрос о том, какое количество MFCC является достаточным для описания одного квазистационарного фрагмента голосового сигнала. При этом в теоретических работах указывается, что правильность выбора вида и параметров нейросетевой модели является важнейшим фактором, предопределяющим эффективность ее применения.

Цель статьи

Основной целью публикации является определение вида и параметров нейросетевой модели, предназначенной для распознавания эмоций диктора на фиксированных участках голосового сигнала.

Изложение основного материала

При определении архитектуры нейросетевой модели использованы общеизвестные методы построения нейросетевых анализаторов параметров технических систем, теоретическая база которых обоснована в [1; 8; 15]. В соответствии с указанными теоретическими наработками, первый этап определения нейросетевой архитектуры соотношен с выбором наиболее эффективного вида нейросетевой модели из множества допустимых видов.

Для формирования множества допустимых видов нейросетевых моделей использованы теоретические результаты [1; 8] и данные [7; 14; 18] о современных достижениях в области применения

нейронных сетей для распознавания голосовых сигналов. Используя указанные результаты, сформировано множество допустимых видов нейросетевых моделей, которое можно представить с помощью следующего выражения:

$$K = \{k_1, k_2, k_3, k_4, k_5\}, \quad (11)$$

где K – множество допустимых видов нейросетевых моделей; k_1 – двухслойный перцептрон; k_2 – глубокие нейронные сети с прямым распространением сигнала; k_3 – сверточные нейронные сети; k_4 – рекуррентные нейронные сети типа LSTM; k_5 – нейронные сети типа топографической карты Кохонена.

Для расчета наиболее эффективного вида нейросетевой модели использованы следующие выражения:

$$N_j = \sum_{l=1}^L q_l R_l(k_j), \quad k_j \in K, j = 1, \dots, J, q_l \in Q; \quad (12)$$

$$\max_{N_j} = \{N_1, N_2, \dots, N_J\}, \quad (13)$$

где J – количество допустимых видов нейросетевых моделей; N_j – функция эффективности для j -го вида нейросетевой модели; $q_l = [0..1]$ – весовой коэффициент l -го критерия эффективности; k_j – j -й вид нейросетевой модели; Q – множество весовых коэффициентов; K – множество допустимых видов нейросетевых моделей; L – количество критериев эффективности; $R_l(k_j)$ – значение l -го критерия для j -го вида нейросетевой модели.

Также на основании сформированных требований к нейросетевой модели определено, что при распознавании эмоционального состояния диктора на фиксированных участках голосового сигнала значимые критерии эффективности определяются: R_1 – возможностью обучения на зашумленных данных; R_2 – объемом памяти, требуемым для обучения сети; R_3 – временем обучения сети; R_4 – точностью распознавания; R_5 – временем принятия решения; R_6 – возможностью экстраполяции результатов обучения; R_7 – возможностью интерпретации выхода в виде вероятности; R_8 – апробированностью в задачах распознавания фиксированных фрагментов голосового сигнала. В первом приближении принята равноценность этих критериев ($q_l = 0,125, l = 1..8$).

Значения представленных в табл. 1 критериев эффективности для каждого вида нейронной сети определены экспертным путем с учетом данных [1; 15]. Определение значений множеств K, R_l и Q позволило рассчитать функцию эффективности для каждого из допустимых видов нейросетевых моделей. Для расчетов использованы выражения (11) – (13). Полученные результаты также показаны в табл. 1.

Подставив данные табл. 1 в выражение (11) для каждого допустимого вида нейросетевой модели, определены значения функции эффективности. Полученные значения показаны в табл. 2.

Таблица 1 – Значения показателей эффективности для допустимых видов нейросетевых моделей

Критерий эффективности	Вид нейросетевой модели				
	k_1	k_2	k_3	k_4	k_5
R_1	1	1	1	0,8	0,5
R_2	0,7	0,8	1	0,5	0,4
R_3	1	0,8	0,7	0,8	1
R_4	1	1	1	0,9	0,5
R_5	1	0,9	0,9	0,8	1
R_6	1	1	1	1	0,9
R_7	1	1	1	0,8	0,8
R_8	1	1	0,5	1	0,8

Таблица 2 – Значения функции эффективности для допустимых видов нейросетевых моделей

Вид нейросетевой модели	Значение функции эффективности
Двухслойный перцептрон (k_1)	0,9625
Глубокие нейронные сети с прямым распространением сигнала (k_2)	0,9375
Сверточные нейронные сети (k_3)	0,8875
Рекуррентные нейронные сети типа LSTM (k_4)	0,825
Нейронные сети типа топографической карты Кохонена (k_5)	0,7375

После подстановки данных табл. 2 в (12) определено, что наиболее эффективным видом является двухслойный перцептрон, для которого функция эффективности достигает максимального значения 0,9625. Отметим, что определение двухслойного перцептрона как вида нейросетевой модели, наиболее эффективного в задаче распознавания эмоций на основании анализа фиксированных фрагментов голосового сигнала, подтверждается данными работ [6; 17]. Типовая структура двухслойного перцептрона показана на рис. 1. Ее особенностью является наличие одного слоя скрытых нейронов.

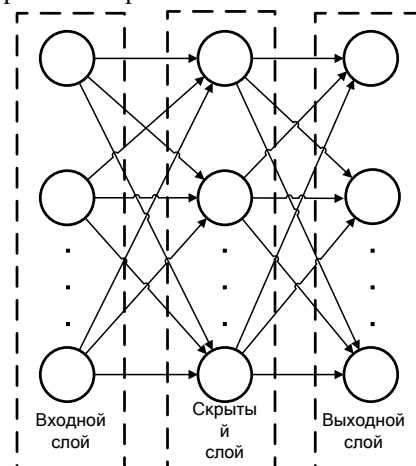


Рисунок 1 – Структура двухслойного перцептрона, предназначенного для распознавания эмоций

Определение наиболее эффективного вида нейросетевой модели позволило перейти к следующему этапу исследований, связанному с определением структурных параметров этой модели. Согласно результатам анализа [1], структурные параметры двухслойного персептрона определяются количеством входных (N_{in}), скрытых (N_h) и выходных нейронов (N_{out}).

По аналогии с [3], в качестве входных параметров двухслойного персептрона предусмотрено использовать мел-кепстральные коэффициенты (MFCC), рассчитанные с помощью выражений (1) – (9) для каждого из квазистационарных фрагментов голосового сигнала. Согласно рекомендациям [3; 6; 9], продолжительность каждого из квазистационарных фрагментов принята равной $t_{st} = 16$ мс. Это объясняется возможностью эффективного применения быстрого преобразования Фурье для расчета спектра квазистационарного фрагмента при частоте дискретизации 16000 Гц, характерной для записей голосовых сигналов в доступных базах.

Отметим, что в проанализированных литературных источниках нет единого мнения о том, какое количество MFCC целесообразно использовать для определения эмоций диктора. В первом приближении принято, что количество MFCC равно 26. Эта цифра принята из соображений простоты реализаций расчета с использованием апробированных программных комплексов. Такое количество MFCC соответствует 22750 входным параметрам нейросетевой модели. При этом расчет количества входных параметров реализован с учетом половинного перекрытия квазистационарных фрагментов при длительности голосового сигнала 7,008 с.

В рассматриваемой задаче количество выходных нейронов равно количеству распознаваемых эмоций, а количество скрытых нейронов можно рассчитать с помощью выражения:

$$N_h = \eta \sqrt{\frac{P}{N_{out}}}, \quad (14)$$

где P – количество учебных примеров, $\eta = [1..10]$ – коэффициент пропорциональности.

Для формирования тренировочной и тестовой выборок использована база данных Emotion_Recognition, доступная на сайте www.kaggle.com. Из представленной базы данных использованы записи голоса одного диктора. Указанный диктор на английском языке в студийных условиях озвучивал 19 текстов разного содержания, проявляя при этом четыре разные эмоции (радость, нейтральность, грусть и злость). Для формирования как учебной, так и тестовой выборки, использовались только первые 7,008 с каждой из записей.

Объем сформированной тренировочной выборки составил 760 примеров, а объем тестовой выборки – 76 примеров. Таким образом, в данной серии экспериментов количество выходных нейронов $N_{out} = 4$, а рассчитанное с помощью выражения (14) количество скрытых нейронов $N_h = 70$.

Модель двухслойного персептрона с сигмоидальной функцией активации скрытых и выходных нейронов реализована программно с использованием языка программирования Python и библиотеки Keras. Также при создании программы использованы библиотеки NumPy – для обработки массивов, Soundfile – для обработки звуковых файлов и Matplotlib – для визуализации результатов.

Эксперименты проводились на персональном компьютере с процессором Intel Core i7-8700 (3.2 – 4.6 ГГц), с объемом оперативной памяти 16 ГБ, видеокартой nVidia GeForce GTX 1660Ti, который функционировал под управлением операционной системы Microsoft Windows 10.

Эффективность двухслойного персептрона оценивалась с помощью показателя точности распознавания (Accuracy) на тренировочной и тестовой выборках при 100 эпохах обучения. На рис. 2 показаны графики зависимости точности распознавания на тренировочной (Train) и тестовой выборках (Test) от количества эпох обучения (Epoch).

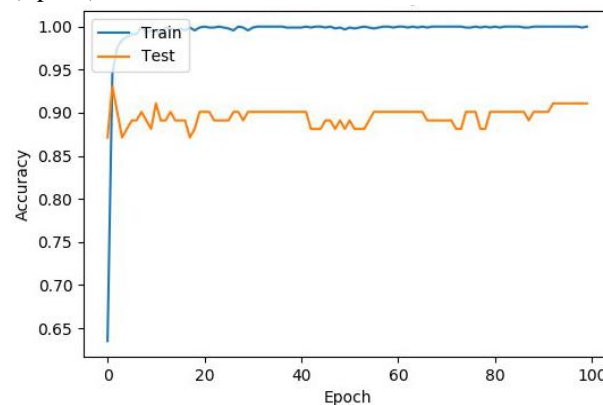


Рисунок 2 – График зависимости точности распознавания от количества эпох обучения для тренировочных и учебных данных

Следует отметить, что достигнутая с помощью двухслойного персептрона точность распознавания эмоций диктора на тестовой выборке составила примерно 0.94. Такой результат соответствует лучшим современным решениям в данной области [2; 5; 14; 18]. Таким образом, результаты экспериментов подтверждают целесообразность распознавания эмоций диктора с помощью разработанного двухслойного персептрона.

Поскольку при проведении экспериментов влияние на эффективность распознавания количества скрытых нейронов, вида функции

активации, а также количества MFCC не рассматривалось, то именно решение этих вопросов и предопределяет пути дальнейших исследований. Также вызывают интерес исследования по нейросетевому распознаванию эмоций диктора в условиях влияния помех различного вида.

Выводы

В результате проведенных исследований обоснована актуальность исследований в направлении разработки средств распознавания эмоций диктора по голосовым фрагментам фиксированной продолжительности. На основании анализа литературного обзора показана перспективность использования нейросетевых решений, создание которых усложняется существующей неопределенностью при выборе вида нейросетевой модели, а также при определении множества входных параметров. В результате

проведенных исследований определено, что в задаче распознавания эмоций диктора по голосовым фрагментам фиксированной продолжительности целесообразно использовать такой вид нейросетевой модели, как двухслойный перцептрон, входные параметры которого ассоциируются с мел-кепстральными коэффициентами, характеризующими каждый из квазистационарных фрагментов анализируемого голосового сигнала. Целесообразность использования двухслойного перцептрона подтверждена с помощью компьютерных экспериментов. Также установлена необходимость проведения дальнейших исследований в направлении адаптации параметров двухслойного перцептрона к условиям распознавания при воздействии помех различного вида. Кроме этого вызывает интерес определение количества мел-кепстральных коэффициентов, которое является достаточным для описания одного квазистационарного фрагмента голосового сигнала.

Список литературы

1. Aitchanov B., Korchenko A., Tereykovskiy I., Bapiyev I. *Perspectives for using classical neural network models and methods of counteracting attacks on network resources of information systems. News of the national academy of sciences of the republic of Kazakhstan series of geology and technical sciences.* 2017. Vol. 5, No 425. Pp. 202-212.
2. Ajinkya N., Nagaraj V., Dharwadkar P. *A Speaker Recognition System Using Gaussian Mixture Model, EM Algorithm and K-Means Clustering. International Journal of Modern Education and Computer Science.* 2018. Vol.10. No.11. Pp. 19-28.
3. Akhmetov B., Tereykovsky I., Doszhanova A., Tereykovskaya L. *Determination of input parameters of the neural network model, intended for phoneme recognition of a voice signal in the systems of distance learning. International Journal of Electronics and Telecommunications.* 2018. Vol. 64, No 4. Pp. 425-432.
4. Altincay H., Demirekler M. *Speaker identification by combining multiple classifiers using Dempster–Shafer theory of evidence. Speech Communication.* 2003. Vol. 41, No 4. Pp.531–547.
5. Ehsan V., Xin L., Erik M., Ignacio L., Javier G. *Deep neural networks for small footprint text-dependent speaker verification. In Acoustics, Speech and Signal Processing.* 2014. IEEE International Conference. Pp. 4052–4056.
6. Geeta N., Soni M. *A New Design Approach for Speaker Recognition Using MFCC and VAD. IJIGSP.* 2013. Vol. 5. No 9. Pp.43 – 49.
7. He L. *Stress and Emotion Recognition in Natural Speech in the Work and Family Environments: Ph.D. Dissertation. – RMIT University,* 2010. 197 p.
8. Hu, Z., Tereykovskiy, I., Zorin, Y., Tereykovska, L., Zhibek, A. *Optimization of convolutional neural network structure for biometric authentication by face geometry. Advances in Intelligent Systems and Computing.* 2018. Vol. 754. Pp 567-577.
9. Ing-Jr D., Chih-Ta Y., Yen-Ming H. *Developments of Machine Learning Schemes for Dynamic Time-Wrapping-Based Speech Recognition. Mathematical Problems in Engineering.* 2013. Pp. 56-68.
10. Jung N, Wranke C, Hamburger K, Knauff M. *How emotions affect logical reasoning: evidence from experiments with mood-manipulated participants, spider phobics, and people with exam anxiety. Front Psychol.* 2014. 5:570.
11. Karam Z., Campbell W. *A new kernel for SVM MLLR based speaker recognition. In: Proc. Interspeech 2007 (ICSLP), Antwerp, Belgium, August 2007. Pp. 290–293.*
12. Konar A., Chakraborty A. *Emotion recognition: a pattern analysis approach. Wiley.* 2015. P. 583.
13. Littlewort, G., Whitehill J., Wu T., Fasel I., Frank M., Movellan J., Bartlett M. *The Computer Expression Recognition Toolbox (CERT). Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition.* 2011, 298 – 305.
14. Satyanand S., Abhay K., David R. *Efficient Modelling Technique based Speaker Recognition under Limited Speech Data. International Journal of Image, Graphics and Signal Processing.* 2016. Vol.8. No.11. Pp.41-48.
15. Tereykovska L., Tereykovskiy I., Mussiraliyeva S., Akhmed G., Beketova A., Sambetbayeva A. *Recognition of emotions by facial Geometry using a capsule neural network. International Journal of Civil Engineering and Technology.* 2019 Vol. 10. Issue 04. Pp. 270-279.
16. Russell J., Bachorowski J., Fernandez-Dols J. *Facial and vocal expressions of emotion. Annu Rev Psychol.* 2003. Vol. 54. Pp. 329-339.
17. Um E., Plass J., Hayward E., Homer B. *Emotional design in multimedia learning. Educ. Psychol. J.* 2012. Pp. 485–498.
18. Zhang W., Deng Y., He L., Liu J. *Variant Time-Frequency Cepstral Features for Speaker Recognition. Interspeech.* 2010. Pp. 2122-2125.

Статья поступила в редакцию 10.02.2020

Терейковська Людмила Олексіївна

Кандидат технічних наук, доцент інформаційних технологій проектування та прикладної математики,
orcid.org/0000-0002-8830-0790

Київський національний університет будівництва і архітектури, Київ

НЕЙРОМЕРЕЖЕВА МОДЕЛЬ РОЗПІЗНАВАННЯ ЕМОЦІЙ ЗА ГОЛОСОМ

Анотація. Розроблено засоби розпізнавання емоційного стану диктора. Показано перспективність використання нейронних мереж для аналізу фіксованих фрагментів голосового сигналу. Встановлено необхідність адаптації виду і параметрів нейромережевої моделі до умов задачі розпізнавання емоцій за голосом. В результаті проведених досліджень визначено, що в задачі розпізнавання емоцій диктора за голосовими фрагментами фіксованої тривалості доцільно використовувати двошаровий перцептрон, вхідні параметри якого асоціюються з мел-кепстральними коефіцієнтами, що характеризують кожен з квазістаціонарних фрагментів голосового сигналу, а вихідні параметри відповідають емоціям диктора, що мають бути розпізнані. Доцільність використання двошарового перцептрону підтверджена за допомогою комп'ютерних експериментів. Визначено, що напрями подальших досліджень пов'язані з визначенням кількості мел-кепстральних коефіцієнтів, що є достатньою для опису одного квазістаціонарного фрагменту, і адаптацією параметрів двошарового перцептрону до умов розпізнавання під впливом різноманітних перешкод.

Ключові слова: емоції; розпізнавання емоцій; голосовий сигнал; нейромережева модель; мел-кепстральні коефіцієнти; квазістаціонарний фрагмент

Tereikovska Liudmyla

PhD (Eng.), Associate Professor, Associate Professor of the Department of Information Technologies of Design and Applied Mathematics, *orcid.org/0000-0002-8830-0790*

Kyiv National University of Construction and Architecture, Kyiv

NEURAL NETWORK VOICE RECOGNITION MODEL

Abstract. The article is devoted to the development of recognition tools for the emotional state of the speaker. The prospects of using neural networks for the analysis of fixed fragments of a voice signal is shown. The necessity of adapting the appearance and parameters of the neural network model to the conditions of the task of recognizing emotions by voice is established. As a result of the studies, it was determined that in the task of recognizing the speaker's emotions by voice fragments of a fixed duration, it is advisable to use a two-layer perceptron, the input parameters of which are associated with mel-cepstral coefficients characterizing each of the quasi-stationary fragments of the analysed voice signal, and the output parameters correspond to the recognizable emotions of the speaker. The feasibility of using a two-layer perceptron is confirmed by computer experiments. It was determined that the directions of further research are related to determining the number of mel-cepstral coefficients, which is sufficient to describe a single quasistationary fragment, and adapting the parameters of the two-layer perceptron to recognition conditions under the influence of various kinds of interference.

Keywords: emotions; recognition of emotions; voice signal; neural network model; mel-cepstral coefficients; quasistationary fragment

Посилання на публікацію

APA Tereikovska, Liudmyla. (2020). Neural network voice recognition model. *Management of Development of Complex Systems*, (41), 95 – 100, [dx.doi.org/10.32347/2412-9933.2020.41.95-100](https://doi.org/10.32347/2412-9933.2020.41.95-100).

ГОСТ Терейковская Л.А. Нейросетевая модель распознавания эмоций по голосу [Текст] / Л.А. Терейковская // *Управление развитием сложных систем*. – 2020. – № 41. – С. 95 – 100, [dx.doi.org/10.32347/2412-9933.2020.41.95-100](https://doi.org/10.32347/2412-9933.2020.41.95-100).