

*К. А. Мартынов<sup>1\*</sup>, Н. Е. Карпова<sup>1</sup>*

## **Использование аппарата нейронных сетей для оценки разборчивости речи**

<sup>1</sup> Самарский государственный технический университет, г. Самара,  
Российская Федерация  
\* e-mail: martyn987@mail.ru

**Аннотация.** В этой статье представлен всесторонний обзор того, как нейронные сети могут быть использованы для объективной оценки разборчивости речи. В статье объясняются традиционные субъективные методы, используемые для оценки разборчивости речи и связанные с ними ограничения. Вводится концепция нейронных сетей и то, как их можно обучить предсказывать показатели разборчивости речи на основе признаков, извлеченных из речевого сигнала. Описываются две популярные архитектуры нейронных сетей: сверточная нейронная сеть (CNN) и рекуррентная нейронная сеть (RNN), которые успешно использовались для оценки разборчивости речи. Освещаются факторы, которые могут повлиять на производительность системы оценки разборчивости речи на основе нейронных сетей. В целом, статья дает обширную информацию о потенциале нейронных сетей для улучшения оценки разборчивости речи и их применении в различных областях.

**Ключевые слова:** искусственные нейронные сети, глубокие нейронные сети, сверточная нейронная сеть, оценка разборчивости речи

*К. А. Martynov<sup>1\*</sup>, N. E. Karpova<sup>1</sup>*

## **Using the Neural Network Apparatus to Assess Speech Intelligibility**

<sup>1</sup> Samara State Technical University, Samara,  
Russian Federation  
\* e-mail: martyn987@mail.ru

**Annotation.** This article provides a comprehensive overview of how neural networks can be used to objectively assess speech intelligibility. The article explains the traditional subjective methods used to assess speech intelligibility and the associated limitations. The concept of neural networks and how they can be trained to predict speech intelligibility indicators based on features extracted from a speech signal is introduced. Two popular neural network architectures are described: convolutional neural network (CNN) and recurrent neural network (RNN), which have been successfully used to assess speech intelligibility. The factors that can affect the performance of a speech intelligibility assessment system based on neural networks are highlighted. In general, the article provides extensive information about the potential of neural networks to improve the assessment of speech intelligibility and their application in various fields.

**Keywords:** artificial neural networks, deep neural networks, convolutional neural network, speech intelligibility assessment

### ***Введение***

Разборчивость речи является важнейшим аспектом общения, который относится к способности слушателя понимать произносимые слова. Точная оценка

разборчивости речи необходима в различных областях, включая аудиологию, логопедию и телекоммуникации. Традиционно оценка разборчивости речи проводилась с помощью субъективных методов, что может отнимать много времени и приводить к ошибкам [1, 8, 9]. Однако с недавними достижениями в области глубокого обучения нейронные сети стали многообещающим подходом к объективной оценке разборчивости речи. В этой статье рассматривается, как нейронные сети можно использовать для оценки разборчивости речи.

### *Методы машинного обучения*

Нейронные сети – это вычислительные модели, основанные на структуре и функциях человеческого мозга. Они состоят из взаимосвязанных узлов или нейронов, которые выполняют простые операции над входными данными и передают выходные данные следующему слою нейронов. Выходные данные последнего слоя нейронов обеспечивают прогнозирование сети. Процесс обучения нейронной сети включает в себя настройку весов соединений между нейронами, чтобы свести к минимуму разницу между выходными данными сети и базовыми метками истинности.

Чтобы оценить разборчивость речи, нейронную сеть можно обучить на наборе данных образцов речи с соответствующими показателями разборчивости. Набор данных может быть получен с помощью субъективных тестов, в которых слушатели оценивают образцы речи на основе их способности понимать произносимые слова. Нейронная сеть обучена предсказывать оценку разборчивости на основе признаков, извлеченных из речевого сигнала.

Одной из самых популярных архитектур нейронных сетей, используемых для оценки разборчивости речи, является сверточная нейронная сеть (CNN) (рис. 1). CNN обычно используются в задачах обработки изображений, а также показала многообещающие результаты в обработке речи. В CNN входной речевой сигнал сначала преобразуется в спектрограмму, которая представляет собой двумерное представление частотного содержания речевого сигнала с течением времени. Затем CNN учится извлекать соответствующие характеристики из спектрограммы, которые используются для прогнозирования показателя разборчивости [5].

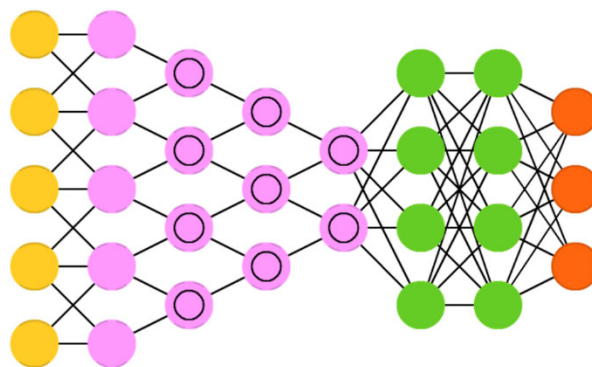


Рис. 1. Сверточная нейронная сеть

Производительность системы оценки разборчивости речи на основе нейронной сети зависит от нескольких факторов, таких как качество обучающего набора данных, выбор архитектуры нейронной сети и метод извлечения признаков. Кроме того, на производительность нейронной сети может влиять фоновый шум и акцент говорящего. Поэтому важно тщательно спроектировать обучающий набор данных и архитектуру нейронной сети для достижения наилучшей производительности.

### *Обучение и достоверность реализованной нейронной сети*

Обучение сверточной нейронной сети происходит на спектрограммах в логарифмическом виде из аудиозаписи мужского голоса с наложенным белым шумом (рис. 2, рис. 3).

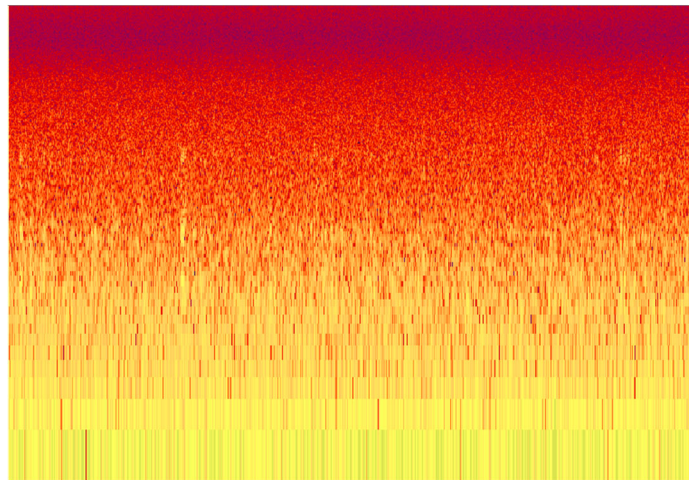


Рис. 2. Спектрограмма из аудиозаписи

```
0.6636
Epoch 4996/5000
110/110 [-----] - 1s 7ms/step - loss: 1.1527 - accuracy: 0.4209 - val_loss: 2.1001 - val_accuracy: 0.8182
Epoch 4997/5000
110/110 [-----] - 1s 7ms/step - loss: 1.2951 - accuracy: 0.4324 - val_loss: 2.1113 - val_accuracy: 0.8636
Epoch 4998/5000
110/110 [-----] - 1s 7ms/step - loss: 1.2793 - accuracy: 0.2818 - val_loss: 1.1710 - val_accuracy: 0.7182
Epoch 4999/5000
110/110 [-----] - 1s 7ms/step - loss: 1.2919 - accuracy: 0.4000 - val_loss: 1.1290 - val_accuracy: 0.9182
Epoch 5000/5000
110/110 [-----] - 1s 7ms/step - loss: 1.1904 - accuracy: 0.4455 - val_loss: 1.8576 - val_accuracy: 0.8818

In [ ]: model.evaluate_generator(generator=test_set, steps=22)
Out[ ]: [1.6221108436584473, 0.27272728085517883]
```

Рис. 3. Точность распознавания разборчивости речи на 4996-5000 эпохе

Проверку достоверности произведем с помощью меры точности (precision) [6]. Результаты процесса обучения введем в TensorBoard, это инструмент для визуализации показателя точности на каждой итерации эпохи (рис. 4), можем увидеть, что средняя достоверность 0,789 [7].

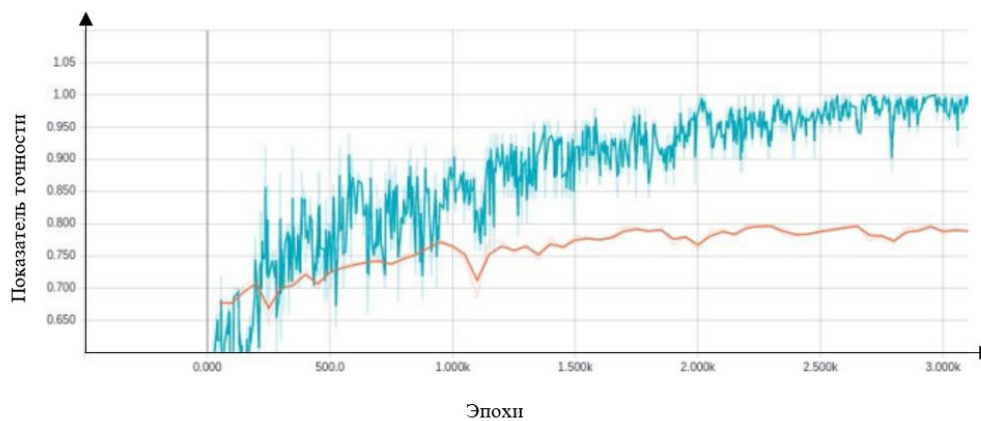


Рис. 4. Доля корректных прогнозов (accuracy): синий график – обучение, оранжевый – проверка

### Заключение

В заключение можно сказать, что нейронные сети продемонстрировали большой потенциал в объективной оценке разборчивости речи. Способность точно оценивать разборчивость речи может иметь значительные последствия в различных областях, включая аудиологию, логопедию и телекоммуникации. С дальнейшими достижениями в области глубокого обучения можно ожидать появления более сложных моделей нейронных сетей, которые могут еще больше повысить эффективность оценки разборчивости речи.

### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Акустическая экспертиза каналов речевой коммуникации. Монография / Дидковский В. С., Дидковская М. В., Продеус А. Н. – Киев, 2008. 420.
2. Чипига, А.Ф. Информационная безопасность автоматизированных систем / А.Ф. Чипига. – М.: Гелиос АРВ, 2017. – 336 с.
3. ОЗИ Практикум ВМ Алефиренко, ЮВ Шамгин, БГУИР 2004 (Лаб практикум).
4. Малюк, А.А. Информационная безопасность: концептуальные и методологические основы защиты информации / А.А. Малюк. – М.: ГЛТ, 2016. – 280 с.
5. J. Turian et al. Word representations: A simple and general method for semi-supervised learning. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, с. 384-394, 2010.
6. В. Д. Чабаненко. Модификации метода стохастического градиентного спуска для задач машинного обучения с большими объемами данных. Master's thesis, Московский государственный университет имени М.В. Ломоносова, 2016.
7. Google Research Team. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv:1603.04467 [cs.DC], 2016.
8. Партыка, Т.Л. Информационная безопасность: учебное пособие / Т.Л. Партыка, И.И. Попов. – М.: Форум, 2016. – 432 с.
9. Петров, С.В. Информационная безопасность: учебное пособие / С.В. Петров, И.П. Слинькова, В.В. Гафнер. – М.: АРТА, 2016. – 296 с.
10. Семенов, В.А. Информационная безопасность: учебное пособие / В.А. Семенов. – М.: МГИУ, 2017. – 277 с.

© К. А. Мартынов, Н. Е. Карпова, 2023