

---

## ЭНЕРГЕТИКИ ОПТИМИЗАЦИЯ ПАРАМЕТРОВ КОДЕКОВ СЖАТИЯ АУДИО НА ОСНОВЕ АНАЛИЗА ВОСПРИЯТИЯ ЧЕЛОВЕКА С ИСПОЛЬЗОВАНИЕМ НЕЙРОННЫХ СЕТЕЙ

**Кирпичев Денис Сергеевич,**

аспирант, Пензенский государственный технологический университет, аспирант  
Россия, г. Пенза  
kirpichev.1999@mail.ru

### Аннотация

---

В данной статье использовалась нейронная сеть для предсказания и реконструкции аудиофайла после сжатия с применением преобразования Фурье. Был проведен упрощенный процесс оптимизации для определения наилучшей структуры нейронной сети с целью максимизации отношения сигнал/шум (SNR). В результате была выбрана сеть с тремя скрытыми слоями. Входными данными для нейросети служил сжатый аудиофайл, а выходными — исходный аудиофайл без сжатия. Обученная сеть смогла восстановить оригинальный сигнал с SNR, равным 1.3946 для левого канала и 1.7734 для правого канала. Реализация была выполнена в среде Matlab.

---

**Ключевые слова:** сжатие аудио, нейронная сеть, оптимизация, преобразование Фурье, отношение сигнал/шум.

---

## OPTIMIZATION OF AUDIO COMPRESSION CODEC PARAMETERS BASED ON HUMAN PERCEPTION ANALYSIS USING NEURAL NETWORKS

**Denis S. Kirpichev,**

Postgraduate student, Penza State Technological University, Postgraduate student  
Russia, Penza  
kirpichev.1999@mail.ru

---

### ABSTRACT

---

In this article, a neural network was used to predict a reconstruct an audio file after a compression using Fourier transformer. A simple optimization process was conducted to find best structure of neural network to maximize Signal to Noise Ratio. a network with three hidden layers were adopted finally, the input of neural network is a compressed audio file and output is the original audio file without compression, the trained network was able to reconstruct the original file with SNR equal to 1.3946 for left channel and 1.7734 for right channel. the implementation was carried out in Matlab environment.

---

**Keywords:** audio compression, Neural Network, Optimization, Fourier transform, Signal to Noise Ratio

Сжатие цифрового аудио обеспечивает эффективное хранение и передачу аудиоданных. Различные методы сжатия аудиосигналов обладают разной степенью сложности, качеством сжатого звука и уровнем уменьшения объема данных [1]. Один из наиболее простых методов сжатия основан на применении преобразования Фурье [2]. Для аудиосигнала  $x$  длины  $n$  преобразование Фурье с числом точек  $L$  может принимать различные формы:

1. Если  $n < L$  преобразованный сигнал  $F_t$  будет иметь длину  $L$ , при этом  $n$  значений будут ненулевыми, а  $(L - n)$  значений окажутся равными нулю.
2. Если  $n > L$ , преобразование Фурье приведет к сжатию сигнала до  $L$  точек в пространстве Фурье.

Параметр  $L$  оказывает непосредственное влияние на процесс сжатия. Обратное преобразование Фурье позволяет восстановить сигнал  $x$  из преобразованного сигнала  $F_t$

Анализ восприятия человеком играет важную роль в области сжатия аудио. Однако данный анализ может варьироваться в зависимости от индивидуальных особенностей восприятия, хотя он остается одним из основных параметров для оценки качества сжатия. Тем не менее, он не является достаточным сам по себе, что требует использования математических методов, таких как отношение сигнал/шум (SNR), как было описано позже.

Цель данного исследования заключается в использовании нейронной сети (НС) для предсказания восстановленного аудиосигнала после сжатия. Предполагается использовать сжатый сигнал для предсказания истинного несжатого сигнала. Это требует оптимизации и выбора наилучших параметров, среди которых:

1. Параметр  $L$  определяющий степень сжатия;
2. Архитектура нейронной сети.
3. Оптимизация осуществляется путем тестирования различных структур нейронной сети, а критерием качества является максимизация отношения сигнал/шум (SNR).

В своей работе A. Défossez и др. [3] авторы представили высокоточный аудиокодек реального времени, использующий нейронные сети. Разработанный метод основан на потоковой архитектуре кодировщика-декодировщика с квантованным скрытым пространством, обученной в сквозном режиме. Для упрощения и ускорения процесса обучения применяется единый многомасштабный спектрограммный противник, который эффективно снижает артефакты и обеспечивает высокое качество получаемых аудиосэмплов. Авторы также предлагают новый механизм балансировки функций потерь, позволяющий стабилизировать процесс обучения: вес отдельной функции потерь, теперь определяет долю общего градиента, которую она должна представлять, тем самым исключая зависимость выбора этого гиперпараметра от масштаба самой функции потерь. Кроме того, в работе исследуется возможность использования компактных моделей трансформеров для дополнительного сжатия полученного представления до 40%, при этом обеспечивая скорость работы, превышающую реальное время.

В своей работе N. M. Khan, G. M. Khan [4] авторы предложили новую модель реконструкции аудиосигнала, использующую нелинейный алгоритм оценки, основанный на Картизианском генетическом программировании с эволюционирующей искусственной нейронной сетью (CGPANN). CGPANN обеспечивает значительно более точное восстановление нелинейных графиков аудиосигналов по сравнению с традиционными методами интерполяции и экстраполяции. Проведено сравнение данных методов по

улучшению отношения сигнал/шум (SNR) и способности работать с зашумленными данными. В отличие от других классических алгоритмов реконструкции, предложенный метод способен восстанавливать сигнал, поврежденный шумом до 50%. В данной работе представлен передовой метод восстановления аудиосигнала с использованием машинного обучения. Эффективность предложенного алгоритма оценивается путем измерения улучшения отношения сигнал/шум (SNR), а также разницы между исходным и восстановленным сигналом с точки зрения средней абсолютной процентной ошибки (MAPE). Зафиксированы следующие улучшения SNR: до 20 дБ при восстановлении одиночных точек с 25% отсутствующих выборок, 19 дБ при многоточечном восстановлении (до 5 точек) при потере 50% данных, а также 16 дБ для сигналов с шумом, имеющим случайное распределение

Рассмотрим монофонический сигнал  $x$  длины  $n$ , сжатый сигнал длины  $L$ . Обозначим  $F_t$  как преобразование Фурье сигнала  $x$  с параметром  $L$ , а  $F_{true}$  – преобразование Фурье исходного сигнала без сжатия ( $L = n$ ) [5-6].

Нейронная сеть получает на входе сжатый сигнал  $F_t$ , а желаемый выходной сигнал –  $F_{true}$ . Пусть  $X_{predicted}$  – восстановленный сигнал, полученный на выходе обученной нейросети. Отношение сигнал/шум (SNR) рассчитывается по следующей формуле [7]:

$$SNR = 20 \cdot \log \left( \frac{x}{x - X_{predicted}} \right)$$

где предполагается, что разница между  $x$  и  $X_{predicted}$  представляет собой шум. В случае стереофонического сигнала каждый канал обрабатывается отдельно, после чего каналы объединяются в восстановленный стереосигнал. Полный процесс представлен ниже, было протестировано множество структур нейронных сетей.

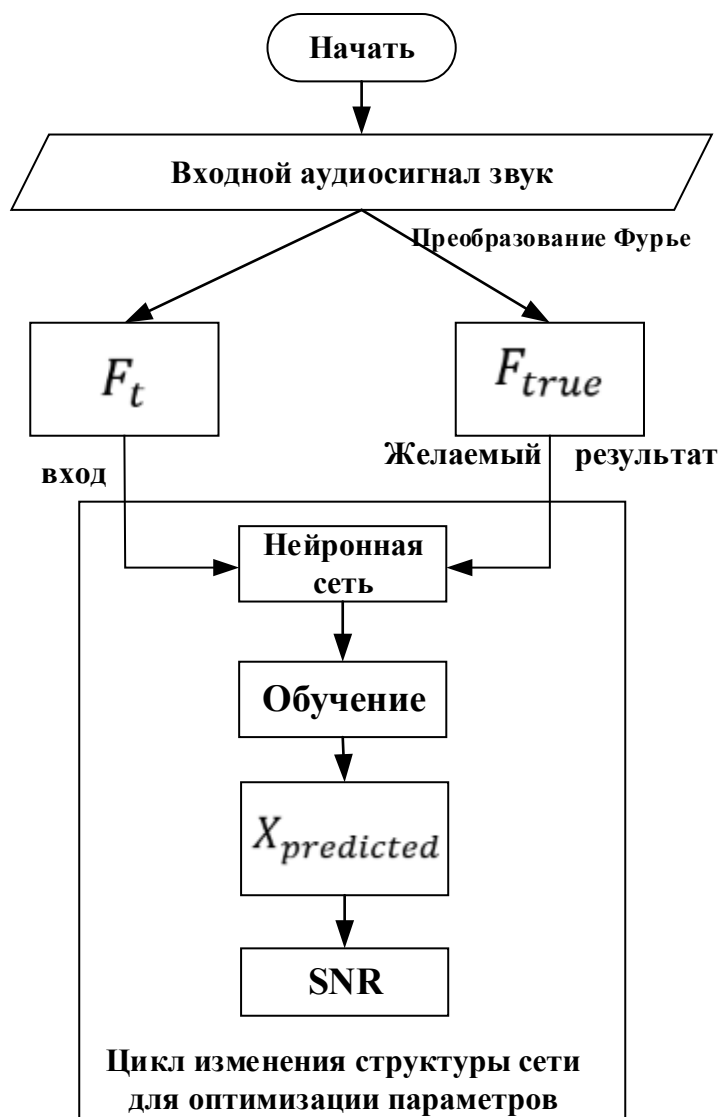


Рисунок 1 - Блок-схема полного процесса

Для эксперимента использовался аудиофайл, полученный из сети. Оптимальная структура нейронной сети имела следующий вид:

Входной слой.

Три скрытых слоя с количеством нейронов [15, 15, 10] соответственно.

Выходной слой.

Производительность процесса обучения представлена на рисунке 2, регрессионная кривая обучения показана на рисунке 3.

Левый и правый каналы исходного аудиосигнала, а также восстановленный аудиофайл приведены на следующих рисунках.

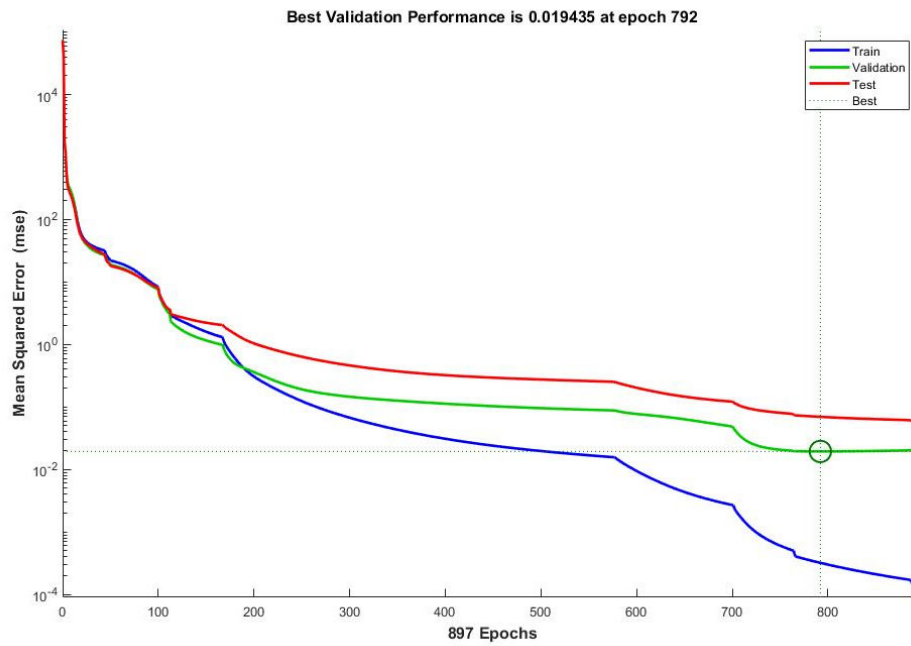


Рисунок 2 - Кривая производительности

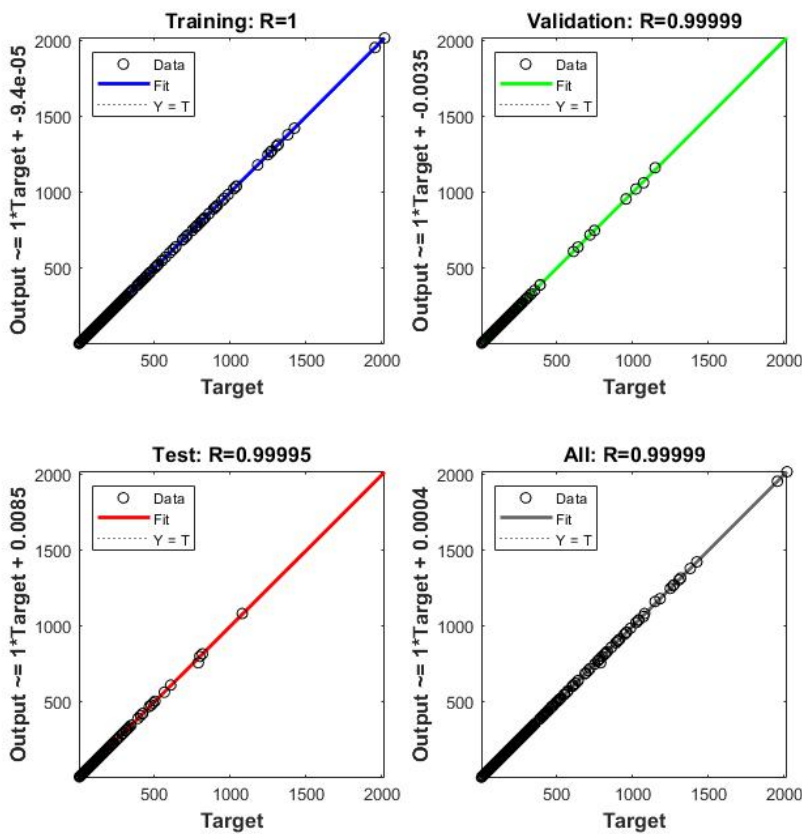


Рисунок 3 - Кривая регрессии

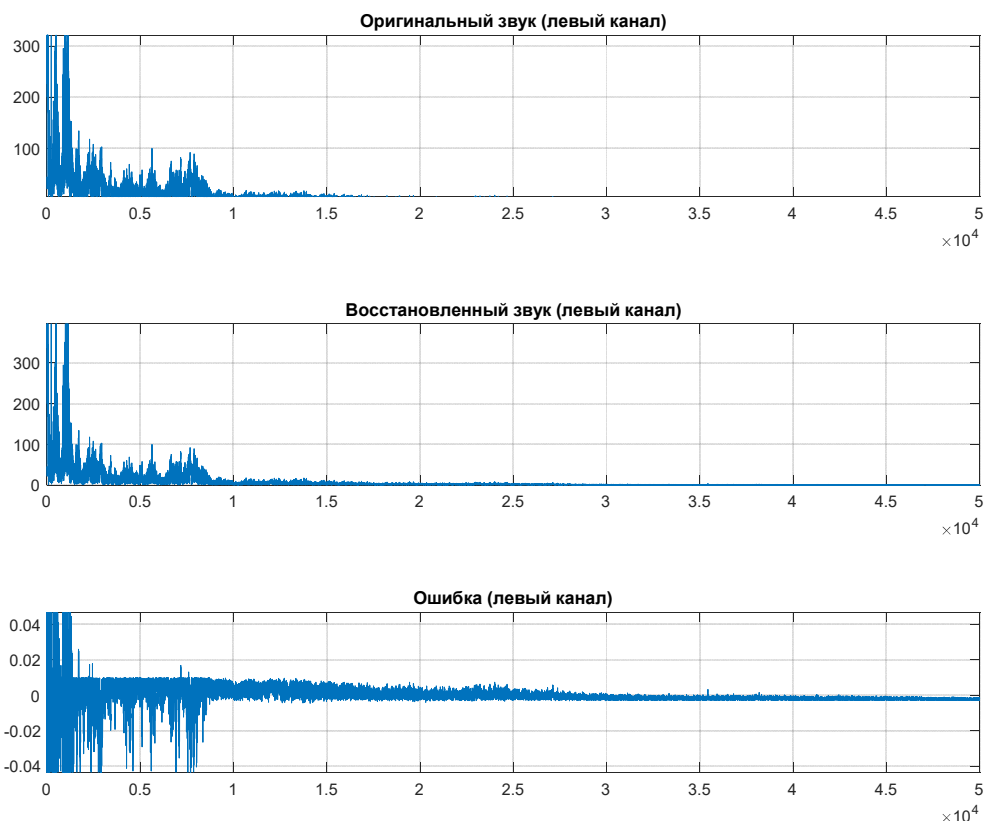


Рисунок 4 - Левый канал

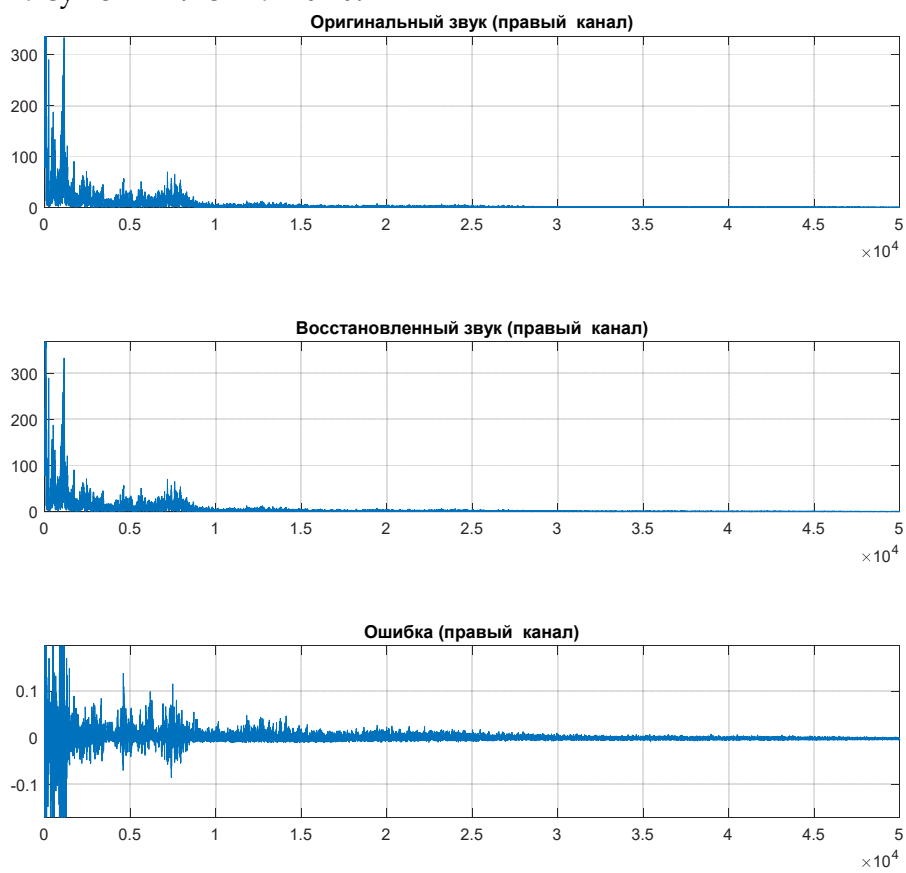


Рисунок 5 - Правый канал

SNR показан на следующем рисунке

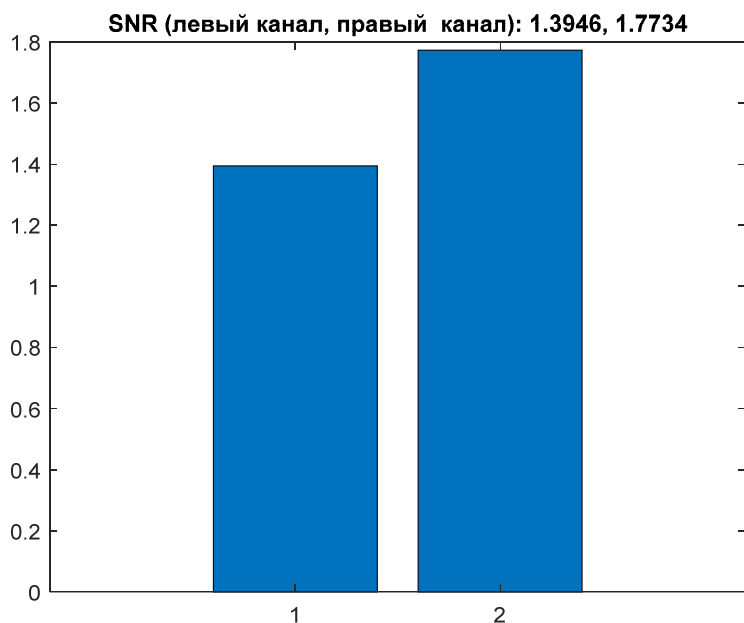


Рисунок 6 - SNR

Результаты показали, что данная архитектура нейронной сети надежна и восстановила аудиофайл на высоком уровне в соответствии с коэффициентом SNR, а ошибка восстановления близка к нулю для обоих каналов. Восстановленный аудиофайл воспроизводится точно также, как исходный.

В заключение, разработанная нейронная сеть обладает способностью восстанавливать аудиофайл после сжатия. Процесс разработки основывался на оптимизационном подходе, направленном на максимизацию параметра отношения сигнал/шум (SNR).

#### Список литературы:

1. Куницын И. А. Устройство сжатия аудиосигнала для передачи по каналам распространения данных. – 2017. – С. 3-5.
2. Мясникова Н. В. и др. Экспресс-анализ сигналов в инженерных задачах. – 2016. – С. 29-30.
3. Défossez A. et al. High-fidelity neural audio compression //arXiv preprint arXiv:2210.13438. – 2022. – С. 1-7.
4. Khan N. M., Khan G. M. Audio signal reconstruction using cartesian genetic programming evolved artificial neural network (CGPANN) //2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA). – IEEE, 2017.
5. Langille P. An Implementation of Audio Compression Using Fourier Analysis: дис. – University of Massachusetts Lowell, 2023. – С. 568-573.
6. Jena S. K. Discrete fourier transform //Fourier, Laplace, and the Tangled Love Affair with Transforms: The Art of Signal Synthesis and Analysis. – Cham: Springer Nature Switzerland, 2025. – С. 229-273.
7. Zhu S. et al. High Signal-to-Noise Ratio MEMS Noise Listener for Ship Noise Detection //Remote Sensing. – 2023. – T. 15. – №. 3. – С. 777.

#### References:

1. Kunitsyn I. A. Audio compression device for transmission over data distribution channels. – 2017. – P. 3-5.

2. Myasnikova N. V. et al. Express analysis of signals in engineering tasks. – 2016. – P. 29-30.
3. Défossez A. et al. High-fidelity neural audio compression //arXiv pre-print arXiv:2210.13438. – 2022.
4. Khan N. M., Khan G. M. Audio signal reconstruction using cartesian genetic programming evolved artificial neural network (CGPANN) //2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA). – IEEE, 2017. – pp. 568-573.
5. Langille P. An Implementation of Audio Compression Using Fourier Analysis: dis. – University of Massachusetts Lowell, 2023. – P. 568-573.
6. Jena S. K. Discrete fourier transform //Fourier, Laplace, and the Tangled Love Affair with Transforms: The Art of Signal Synthesis and Analysis. – Cham: Springer Nature Switzerland, 2025. – P. 229-273.
7. Zhu S. et al. High Signal-to-Noise Ratio MEMS Noise Listener for Ship Noise Detection //Remote Sensing. – 2023. – T. 15. – №. 3. – C. 777.