

DOI: <https://doi.org/10.17816/gc623433>

# Вокодер WaveNet в задаче предсказания временного ряда с экстремальными событиями

Н.В. Громов\*, Т.А. Леванова

Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского, Нижний Новгород, Российская Федерация

## АННОТАЦИЯ

Экстремальными событиями обычно считаются редкие и непредсказуемые события или события, которые сильно отклоняются от типичного поведения. Однако объективные критерии экстремальных событий ещё не определены. Редкость требует некоторых характерных масштабов или временных и пространственных границ, в то время как интенсивность должна отражать способность события вызвать большие изменения. Одним из наиболее ярких примеров экстремальных событий в нейронауке и медицине являются эпилептические припадки [1].

В синтезе речи сети-вокодеры, такие как WaveNet [2], используются для генерации аудиодорожек. Эта модель представляет из себя свёрточную нейронную сеть, представленную как каузальный фильтр, который не смотрит в будущее. Эта особенность вокодера может быть потенциально применима к предсказанию временных рядов. Звуковые временные ряды тоже могут быть рассмотрены как динамическая система, довольно хаотическая, с переключением режимов (например, переход от одной буквы к другой). В таких рядах присутствуют большие отклонения амплитуды, что может быть сравнимо с экстремальным событием. Эта сеть получает  $r$  последних временных отсчётов (так называемое рецептивное поле) на вход и предсказывает следующий за ними отсчёт, основываясь на этом рецептивном поле. Структура сети имеет древовидный вид. Расстояние между входами на последующие слои растёт экспоненциально. Это необходимо, потому что рецептивное поле  $r$  обычно достаточно большое, порядка тысячи или двух тысяч. И если не увеличивать экспоненциально расстояние между входами, то количество слоёв будет линейно зависеть от  $r$ . У рекуррентных нейронных сетей, широко применяемых в задачах предсказания последовательностей, есть следующая проблема: во временных рядах им проще предсказывать очень похожий семпл на последний увиденный для оптимизации функции потерь и тем самым сеть в предсказании может сойтись к моде, а в свёрточной сети выход на моду будет дольше, потому что рецептивное поле достаточно большое: в звуке, например, на нём происходит множество осцилляций и сеть не отдаёт предпочтения какому-либо отсчёту.

Эксперименты были проведены с искусственными данными, сгенерированными системой двух нейронов Хидмарш-Роуз с химической синаптической связью. Мы выбрали суммарный мембранный потенциал как наблюдаемую переменную, основываясь на биологической составляющей этой системы. Описанная система демонстрирует экстремальные события для широкого набора параметров связи. Численный критерий для экстремальных событий выбран аналогично тому, как это было сделано в прошлых исследованиях [3]. Вокодер WaveNet демонстрирует точность 91% и полноту 82% для предсказательной длины, равной ширине экстремального события. Заметим, что полнота в данной задаче более важна, так как она отвечает за те случаи, когда модель предсказывает ложное отсутствие экстремального события.

**Ключевые слова:** экстремальные события; свёрточные нейронные сети; машинное обучение; нейрон Хидмарш-Роуз; WaveNet; речевые технологии.

## Как цитировать:

Громов Н.В., Леванова Т.А. Вокодер WaveNet в задаче предсказания временного ряда с экстремальными событиями // Гены и клетки. 2023. Т. 18, № 4. С. 847–849. DOI: <https://doi.org/10.17816/gc623433>

## ДОПОЛНИТЕЛЬНАЯ ИНФОРМАЦИЯ

**Вклад авторов.** Все авторы подтверждают соответствие своего авторства международным критериям ICMJE (все авторы внесли существенный вклад в разработку концепции, проведение исследования и подготовку статьи, прочли и одобрили финальную версию перед публикацией).

**Источник финансирования.** Работа была поддержана грантом Российского научного фонда № 19-72-10128.

**Конфликт интересов.** Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Рукопись получена: 14.05.2023

Рукопись одобрена: 26.11.2023

Опубликована online: 20.01.2024

## СПИСОК ЛИТЕРАТУРЫ

1. Engel Jr.J., Pedley T.A. Generalized convulsive seizures. In: Tassinar C.A., Michelucci R., Shigematsu H., et al., editors. *Epilepsy: a comprehensive text-book*. 1997.
2. Van den Oord A., Dieleman D., Zen H., et al. Wavenet: a generative model for raw audio // arXiv. Vol. 1609. P. 03499. doi: 10.48550/arXiv.1609.03499
3. Gromov N., Gubina E., Levanova T. Loss functions in the prediction of extreme events and chaotic dynamics using machine learning approach. In: *Proceedings of the Fourth International Conference Neurotechnologies and Neurointerfaces (CNN); 2022 Sept 14–16; Kaliningrad, Russian Federation. Kaliningrad, 2022. P. 46–50. doi: 10.1109/CNN56452.2022.9912515*

## КОНТАКТНАЯ ИНФОРМАЦИЯ

\* Н.В. Громов; адрес: Российская Федерация, 603022, Нижний Новгород, пр-т Гагарина, д. 23; e-mail: gromov@itmm.unn.ru

DOI: <https://doi.org/10.17816/gc623433>

# WaveNet vocoder for prediction of time series with extreme events

N.V. Gromov\*, T.A. Levanova

National Research Lobachevsky State University of Nizhny Novgorod, Russian Federation, Nizhny Novgorod, Russian Federation

## ABSTRACT

Extreme events are typically defined as rare or unpredictable events that deviate significantly from typical behavior. Despite this, objective criteria for extreme events have yet to be established. Rareness may be characterized by certain scales or spatial and temporal boundaries, while intensity is an indication of an event's potential to cause a significant change. One of the most prominent occurrences of extreme events in both neuroscience and medicine is in the case of epileptic seizures [1].

In speech synthesis, vocoder networks like WaveNet [2] generate audio. The model is a multi-layer convolutional neural network that functions as a causal filter and doesn't predict the future. Due to this quality, the vocoder may have potential in time series prediction. Audio time series can be regarded as a dynamic system characterized by unpredictable switching regimes. For instance, transitioning from one letter to another can result in significant deviations in amplitude, similar to extreme events. This network receives  $r$  previous input counts known as a receptive field, and uses them to predict the next sample. The network is tree-like in structure, with exponentially increasing distances between subsequent layers of inputs. This is a necessary feature since the receptive field  $r$  is usually quite large, on the order of one or two thousand. Without this exponential increase in distance, the number of layers would depend linearly on  $r$ . Recurrent neural networks pose a challenge in optimizing the loss function when predicting time series sequences, as they tend to predict samples very similar to the previous one, causing the network to converge towards the mode. However, in a convolutional network, the output to the model will be longer due to the large receptive field. In the case of sound analysis, for instance, multiple oscillations occur within a given timeframe and the network does not elevate any specific sample.

The study used artificial data generated from two coupled Hidmarsh–Rose neurons with chemical synaptic couplings. The observed variable was determined by the biological significance of the system, specifically the total membrane potential. The results exhibited extreme events across various coupling parameter values. Based on prior research [3], a numerical standard was selected for the events. The WaveNet vocoder model exhibits a 91% accuracy rate and 82% recall rate when forecasting extreme events of the same width as the prediction. It is noteworthy that recall is crucial in the forecast of extreme events since it identifies instances where the model predicted falsely that an extreme event would not occur.

**Keywords:** extreme events; convolutional neural networks; machine learning; Hidmarsh–Rose neuron; WaveNet; speech technologies.

## To cite this article:

Gromov NV, Levanova TA. WaveNet vocoder for prediction of time series with extreme events. *Genes & cells*. 2023;18(4):847–849. DOI: <https://doi.org/10.17816/gc623433>

## ADDITIONAL INFORMATION

**Authors' contribution.** All authors made a substantial contribution to the conception of the work, acquisition, analysis, interpretation of data for the work, drafting and revising the work, final approval of the version to be published and agree to be accountable for all aspects of the work.

**Funding sources.** This study was supported by the Russian Science Foundation grant No. 19-72-10128.

**Competing interests.** The authors declare that they have no competing interests.

## REFERENCES

1. Engel JrJ, Pedley TA. Generalized convulsive seizures. In: Tassinari CA, Michelucci R, Shigematsu H, et al, editors. *Epilepsy: a comprehensive text-book*. 1997.
2. Van den Oord A, Dieleman S, Zen H, et al. Wavenet: a generative model for raw audio. *arXiv*. 2016;1609.03499. doi: 10.48550/arXiv.1609.03499
3. Gromov N, Gubina E, Levanova T. Loss functions in the prediction of extreme events and chaotic dynamics using machine learning approach. In: *Proceedings of the Fourth International Conference Neurotechnologies and Neurointerfaces (CNN)*; 2022 Sept 14–16; Kaliningrad, Russian Federation. Kaliningrad; 2022. P. 46–50. doi: 10.1109/CNN56452.2022.9912515

## AUTHORS' CONTACT INFO

\* N.V. Gromov; address: 23 Gagarin avenue, 603022 Nizhny Novgorod, Russian Federation; e-mail: [gromov@itmm.unn.ru](mailto:gromov@itmm.unn.ru)

**Received:** 14.05.2023

**Accepted:** 26.11.2023

**Published online:** 20.01.2024