

Таблица 1 – Среднее значение параметров

Параметр	Среднее значение у здорового, %	Среднее значение с патологией, %
$Jitter_{loc}$	0.5185	0.5838
$Jitter_{rap}$	0.3057	0.3616
$Jitter_{ppq5}$	0.3010	0.3611
$Shimmer_{loc}$	4.0631	7.0135
$Shimmer_{app3}$	2.2595	3.8668
$Shimmer_{app5}$	2.4149	4.2968
$Shimmer_{app11}$	2.9780	5.2461

Проанализировав графики и таблицу можно заметить, что для голоса с патологией характерны более высокие показания джиттера и шиммера.

**Список использованных источников:**

1. The dysphonia severity index: an objective measure of vocal quality based on a multiparameter approach / F.L. Wuyts, M.S. De Bodt, G. Molenberghs et al. // *Speech Lang Hear Res.* — 2000. — Vol. 43(3). — P. 796–809.
2. Damste P.H. Disorders of voice // *Scott Brown's Otolaryngology.* — 6-th ed. — *Laryngology and Head and Neck Surgery.* — Oxford: ButterworthHeinemann, 1997. — Vol. 5. — Ch. 6. — P. 5/6/1–5/6/25.
3. Pedersen M. Vocal folds nodules: a model for a prospectiveblinded randomized study with control groups and adequate follow-up // *8th international congress of pediatric otorhinolaryngology.* — Oxford, 2002. — P. 198.
4. Benda A., Jilbab A., Hammouch A. Voice assessments for detecting patients with neurological diseases using PCA and NPCA // *Int. J. of Speech Technology.* 2017. Vol. 20. P. 673–683.
5. M.I. Vashkevich, A.D. Gvozdoovich, Y.N. Rushkevich, A.A. Petrovsky Acoustic analysis of voice for detection of speech disorder for amyotrophic lateral sclerosis // *Doklady BGUIR.* 2018, Vol. 117, No. 7, pp. 64-68.
6. Titze I.R. Comparison of F0 extraction Methods for High-Precision Voice Perturbation Measurements // *J. of Speech and Hearing Research,* Vol. 36, 1120-1133, December 1993

## УПРАВЛЕНИЕ ЧАСТОТОЙ ОСНОВНОГО ТОНА ГОЛОСА В ЗАДАЧАХ СИНТЕЗА РЕЧИ ПО ТЕКСТУ

*Вашкевич Г.С.*

*Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь*

*Азаров И.С. – д.т.н., доцент*

В данной работе предлагается способ управления частотой основного тона в задаче синтеза речи по тексту. Предложенная модель синтеза основана на популярной модели синтеза речи Tacotron 2. В работе рассматриваются такие вопросы, как обзор современных подходов к модулированию стиля синтезированной речи, подготовка данных для обучения модели, модификация базовой архитектуры Tacotron 2.

Несмотря на то, что на текущий момент системы автоматического синтеза речи способны генерировать речь с достаточно высоким качеством, остается нерешенной проблема управления стилем синтезируемой речи. На данный момент последние достижения этой области демонстрируют неплохие результаты в переносе стиля речи из существующей аудиозаписи на синтезируемую [1, 2]. Недостатком данных подходов является то, что они способны хорошо копировать стиль речи из базовой аудиозаписи только в том случае, если структура синтезированной фразы сильно отличается от базовой. Более того, существующие модели не дают возможности управлять интонацией речи на уровне отдельных слов в синтезируемой фразе, и требуют наличия базовой аудиозаписи, стиль речи из которой будет переноситься в синтезируемую фразу.

Изменение частоты основного тона на протяжении всей фразы, или её отдельных слов частично определяет интонацию речи. В данной работе предлагается подход к модулированию частоты основного тона на уровне символов в синтезируемой фразе. Так, каждому символу в синтезируемом тексте ставится в соответствие одно из трех значений – -1, 0, или 1, определяющих

поведение частоты основного тона. Значение -1 соответствует понижению частоты основного тона при озвучивании соответствующего символа, 0 соответствует отсутствию изменения частоты основного тона, а значение 1 соответствует повышению основного тона при озвучивании соответствующих символов.

В качестве базовой модели искусственной нейронной сети для синтеза речи мы использовали модель Tacotron 2 [3]. Её модификация состоит в том, что кроме признаков, представляющих собой синтезируемый текст, ей на отдельный вход подаются так же признаки, соответствующие изменен частоте основного тона. Предложенная архитектура модели синтеза спектрограммы представлена на рисунке (рисунок 1).



Рисунок 1 – Подпись рисунка отделяется от следующего абзаца пробельной строкой

В качестве вокодера для преобразования мелспектрограммы в аудио использовалась модель FloWaveNet [4], преимуществом которой перед моделью Parallel WaveNet (которая использовалась в [3]), является простота обучения.

Для обучения модели использовался набор данных, объемом около 25 часов русской речи, записанной профессиональным диктором.

Таким образом, основными преимуществами предложенной системы синтеза речи являются: возможность модулирования интонации на уровне отдельных слов, или даже словосочетаний, эффективность обучения вокодера, при примерно одинаковом качестве синтезированной речи.

**Список использованных источников:**

1. Style Tokens: Unsupervised Style Modeling, Control and Transfer in End-to-End Speech Synthesis / Y. Wang [et al.] // arXiv preprint, arXiv:1803.09017v1. – 2018. – 11p.
2. Towards End-to-End Prosody Transfer for Expressive Speech Synthesis with Tacotron / RJ Skerry-Ryan [et al.] // arXiv preprint, arXiv: 1803.09047. – 2018. – 11p.
3. Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions / J. Shen [et al.] // arXiv preprint, arXiv1712.05884v2. – 2018. – 5p.
4. FloWaveNet : A Generative Flow for Raw Audio / K. Sungwon [et al.] // arXiv preprint, arXiv1811.02155v2. – 2018. – 6p.

## **АНАЛИЗ КЛАССИФИКАТОРОВ ДЛЯ ДИАГНОСТИРОВАНИЯ НЕВРОЛОГИЧЕСКОГО ЗАБОЛЕВАНИЯ БАС**

*Гвоздович А.Д.*

*Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь*

*Вашкевич М.И. – к.т.н., доцент*

Подход к диагностированию БАС на основе анализа речи является одним из перспективных направлений в детектировании данного заболевания, не имеющего характерных биомаркеров. На основе выделенных из речевого сигнала признаков необходимо произвести классификацию на здоровых и больных. В данной работе рассматриваются три различных алгоритма классификации и сравниваются между собой.

Боковой амиотрофический склероз (БАС) является прогрессирующим нейродегенеративным заболеванием, поражающим нервные клетки головного и спинного мозга. Медиана выживаемости при БАС составляет 3 года.

Отсутствие характерных биомаркеров БАС приводит к задержке в постановке правильного диагноза. В последнее время набирает популярность подход к детектированию заболевания основанный на анализе речи пациента. Согласно идее, предложенной в [1], для обнаружения характерных речевых нарушений может быть использована только часть речевого теста, содержащая гласные. В [2] сделан анализ и выбор подходящих гласных. Было показано, что гласные /и/ и /а/