

4. Нуссбаумер Г. Быстрое преобразование Фурье и алгоритмы вычисления свертки. — М.: Радио и связь, 1985.
 5. Лайонс Р. Цифровая обработка сигналов / Р. Лайонс; пер. с англ. А.А. Бритова. — 2-е изд. — М.: БИНОМ, 2007.

ОЦЕНКА ПОКАЗАТЕЛЕЙ ПЕРТУРБАЦИИ КВАЗИПЕРИОДИЧЕСКОГО СИГНАЛА В ЗАДАЧАХ ДИАГНОСТИКИ ГОЛОСА

Бурак А.А.

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

Вашкевич М.И. — к.т.н., доцент

Рассматривается способ акустического анализа голосового сигнала, содержащий протяжный звук /а/, для определения статистической зависимости параметров джиттер и шиммер от наличия патологии в речевом аппарате.

В начале 1970-х был сконструирован и внедрен в фоноатрическую практику воксфункциограф, позволяющий анализировать интенсивность звукового давления и частоту основного тона голоса. В дальнейшем данный метод получил название акустический анализ голоса (ААГ). В [1] была разработана методика автоматизированного определения голосового поля. Показатели полученные при помощи ААГ, позволили, с одной стороны, исследователям более детально описывать голосовое поле в норме и при патологии, а с другой, — дали возможность осуществлять контроль за эффективностью проводимого лечения голосовых расстройств [2]. Однако предложенные методы были весьма неудобными и крайне длительными как для врача, так и для испытуемого. В силу этих причин с развитием вычислительной техники были предложены методы акустического анализа с использованием компьютерной обработки данных [3]. Это дало ряд существенных преимуществ. В частности, стало возможным получить «среднюю фонетогрмму» от нескольких испытуемых, что было необходимо для статистической обработки информации.

Важными параметрами для голосового сигнала, полученного в процессе диагностики, являются джиттер и шиммер.

Джиттер — мера пертурбации (возмущений) частоты основного тона, показывающая произвольные изменения в частоте смежных вибрационных циклов голосовых складок [4].

Шиммер — мера аналогичная джиттеру, только характеризующая пертурбации амплитуд сигнала на смежных циклах колебаний основного тона.

Существует несколько вариантов оценок параметров джиттер и шиммер. Так, локальный джиттер определяется выражением

$$Jitter_{loc} = \frac{1}{N-1} \sum_{i=1}^{N-1} |T_i - T_{i+1}| \bigg/ \frac{1}{N} \sum_{i=1}^N T_i \quad (1)$$

где T_i — длительность i -го периода основного тона, а N — число периодов основного тона исследуемого сигнала. Часто для оценки джиттера используют отклонение текущей длительности периода не от предыдущей, а от локально усредненного значения, которое рассчитывается на окне в 3 или 5 выборок [5]:

$$Jitter_{rap} = \frac{1}{N-2} \sum_{i=2}^{N-1} \left| T_i - \left(\frac{1}{3} \sum_{n=i-1}^{i+1} T_n \right) \right| \bigg/ \frac{1}{N} \sum_{i=1}^N T_i \quad (2)$$

$$Jitter_{ppq5} = \frac{1}{N-4} \sum_{i=3}^{N-2} \left| T_i - \left(\frac{1}{5} \sum_{n=i-2}^{i+2} T_n \right) \right| \bigg/ \frac{1}{N} \sum_{i=1}^N T_i \quad (3)$$

Значение параметра шиммер определяется по аналогии с джиттером, только вместо значений длины периодов T_i берутся значения амплитуды колебания на данном периоде A_i . Таким образом, формула для локального шиммера $Shimmer_{loc}$ получается из (1) путем замены T_i на A_i . По аналогии с формулами (2) и (3) существуют варианты, определяющие шиммер, как относительное отклонение амплитуды от локально-усредненного значения на интервале в L выборок:

$$Shimmer_{apqL} = \frac{1}{N-L+1} \sum_{i=1+(L-1)/2}^{N-(L-1)/2} \left| A_i - \left(\frac{1}{L} \sum_{n=i-(L-1)/2}^{i+(L-1)/2} A_n \right) \right| \bigg/ \frac{1}{N} \sum_{i=1}^N A_i$$

Чаще всего L принимает значения 3, 5 и 11.

Для расчета параметров джиттер и шиммер требуется первоначально сегментировать голосовой сигнал на периоды основного тона. В работе использовался метод, являющийся частным случаем метода, описанного [6], коэффициент масштабирования амплитуды не используются для вычисления среднего квадратического отклонения между соседними сегментами. На первом этапе сегментации производится преобразование Фурье. После этого необходимо найти первую гармонику основного тона, поиск производится не на всей частотной области, а в диапазоне 60 - 440 Гц, это обусловлено тем, что ЧОТ взрослого человека лежит в данном интервале. На следующем этапе берется обратное преобразование Фурье от всех гармоник, лежащих в интервале от 60 Гц до первой гармоники основного тона (включительно). В получившемся сигнале определяются грубые метки такие, как пики сигнала и его пересечения с нулем (из отрицательной области в положительную). За начало первого отсчета принимается 1-й найденный пик. Для грубой оценки конца интервала принимается сумма 2-го пересечения с нулем и разницы между 1-м пиком и пересечением с нулем. Далее в некотором интервале от получившегося значения производится поиск такой точки, чтобы СКО между текущим интервалом и следующим было минимально. Эта процедура повторяется до тех пор, пока не будет полностью сегментирован весь сигнал. Предложенный способ сегментации имеет недостатки, одним из которых является то, что берется преобразование Фурье от всего сигнала, т.к. помеха, содержащаяся на каком-то интервале сигнала, может неблагоприятно сказаться на этапе поиска грубых меток, поэтому алгоритм сегментации требует дальнейшего усовершенствования.

Для анализа параметров джиттера и шиммера у людей с и без патологии в речевом аппарате была обработана база записей голосов (норма + патология) с использованием описанной выше методики. Запись голосов производилась в РНПЦ Отоларингологии (г. Минск) под контролем врача фониатра.

По результатам анализа были вычислены статистические характеристики полученных параметров голоса. В качестве примера на рис. 1 показаны диаграммы размаха («ящичковые диаграммы»), полученные для параметров $Jitter_{ppq5}$, $Shimmer_{apq11}$. В таблице 1 приведены средние значения по всем выделенным параметрам для здоровых голосов и голосов с патологией.

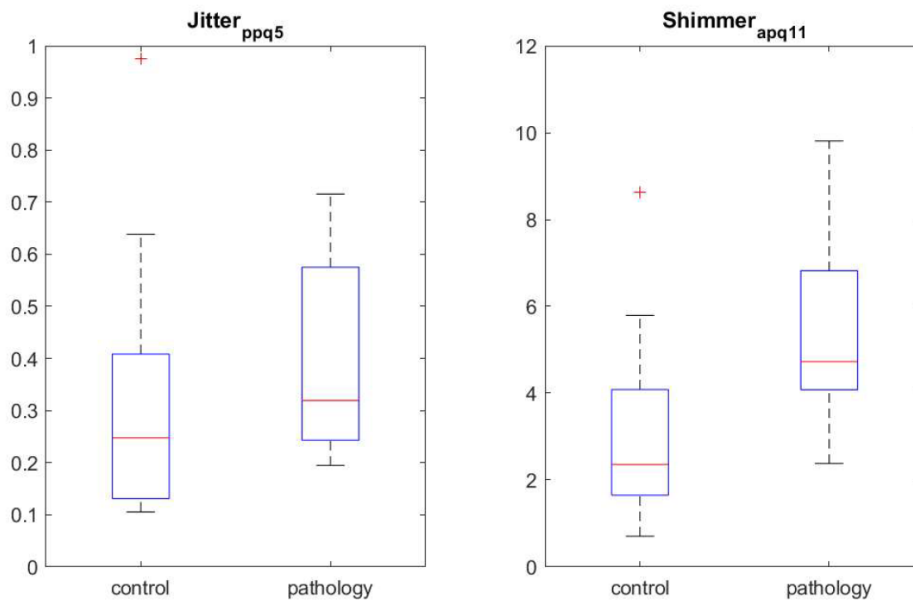


Рисунок 1 – Диаграммы размаха для $Jitter_{ppq5}$, $Shimmer_{apq11}$

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
$Shimmer_{ppq3}$	2.2595	3.8668
$Shimmer_{ppq5}$	2.4149	4.2968
$Shimmer_{ppq11}$	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. Gvozdovich, 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