

# Акустический анализ голоса в норме и патологии

*Лысак А.П., аспирант Института проблем передачи информации РАН*

Выявление заболеваний на ранней стадии является одним из ключевых факторов успешности их лечения, особенно в случаях, связанных с социально важными органами. Акустический анализ голоса с целью определения текущего состояния гортани и классификации «норма/патология» применяется врачами-специалистами на протяжении нескольких лет. Основными критериями при этом служат параметры, характеризующие «Shimmer» и «Jitter» эффекты, уровень турбулентного шума и нелинейность процесса колебаний. Результаты такой классификации показывают ошибку в 5%–20% на речевых корпусах с выраженными заболеваниями. Используемые параметры лишь косвенно описывают текущее состояние голосовых складок, однако, применение методов, позволяющих получить более физиологичные параметры, может увеличить вероятность обнаружения болезни на более ранних стадиях, когда заболевание незначительно влияет на голос человека.

• акустический анализ голоса • патология • речевой сигнал • колебания голосовых складок • «Jitter» эффект • «Shimmer» эффект • турбулентный шум • нелинейный анализ • площадь голосовой щели

Detection of diseases at early stage is the key factor for their successful treatment, especially for diseases of socio important organs. During several years, speech pathologists have been using an acoustic voice analysis that helps them to determine a current state of patient's larynx. Shimmer, jitter, level of turbulent noise and nonlinear features are main criteria for classification of normal/pathological voice. Classification error varies from 5% to 20% for apparent voice disorders; utilized parameters describe current state of vocal cords indirectly. However, usage of more physiological parameters might increase a probability of disease detection on early stage when its influence on person's voice is negligible

• acoustic voice analysis • speech pathology • vocal folds • jitter • shimmer • turbulent noise • nonlinear analysis • square of glottis

## Введение и актуальность задачи

В 2008 г. было диагностировано более 150 тыс. новых случаев рака гортани: более 8 тыс. — на территории РФ, более 12 тыс. — США и более 28 тыс. в странах Евросоюза. Рак гортани стал причиной более чем 80 тыс. смертей [1]. Несмотря на постоянное совершенствование методов диагностики, остаётся стабильной поздняя диагностика заболевания (60–70% составляют III–IV стадии заболевания). В течение первого года после установления диагноза рака гортани умирают 32,8% больных [2].

Существует множество других заболеваний, влияющих на голос. Около 5% населения имеют различные проблемы с голосом [3]. К группе риска относятся: курильщики, педагоги детских садов, преподаватели школ и вузов, экскурсоводы, переводчики, диспетчеры, дикторы и т.д. При выявлении ранней стадии заболевания значительно повышается вероятность его излечения и сокращается период восстановления голосовой функции. Это особенно важно в случае рака гортани, где медикаментозные методы и микрохирургия позволяют сохранить голосовой аппарат и предотвратить распространение злокачественной опухоли.

Нежелание пациентов проходить профилактические осмотры, ограниченное число врачей-специалистов не позволяют выполнять регулярную диспансеризацию с целью выявления ранних стадий патологии. Поэтому разработка средств, сравнимых по простоте применения с флюорографией, или даже более удобных для врачей и пациентов, является актуальной задачей.

Попытки ранней диагностики заболеваний по голосу с использованием компьютерных методов акустического анализа речевого сигнала предпринимаются в течение нескольких десятилетий [4, 5]. Результаты лабораторных исследований показывают, что вероятность правильного отличия нормы от патологии достигает 90–95%. Однако, как это следует из опыта применения систем распознавания речи, устойчивость таких сравнительно высоких оценок в реальных условиях эксплуатации подвергается сомнению.

### Механизм голосообразования

Истинные голосовые складки — две симметрично расположенные складки слизистой оболочки гортани, имеющие особое мышечное строение [6] и выступающие в её полость. При обычном дыхании голосовые складки разведены и образуют голосовую щель в форме равнобедренного треугольника (рис. 1). При фонации истинные голосовые складки находятся в сомкнутом состоянии и совершают колебания благодаря струе выдыхаемого воздуха [7].

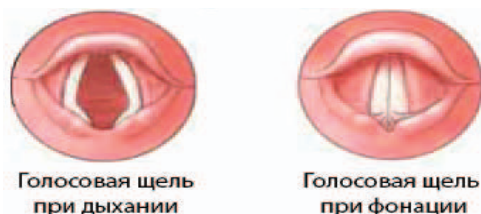


Рис. 1. Рабочие состояния голосовых складок

### Колебания голосовых складок во фронтальной плоскости

Колебания голосовых складок во фронтальной области разделяются на несколько этапов (рис. 2). Увеличение подскладочного давления приводит к постепенному открытию голосовой щели. Проходящая через голосовую щель струя воздуха способствует созданию эффекта Бернулли и появлению турбулентных шумов в надскладочной области, а совместные действия силы упругости и аэродинамической силы приводят к тому, что складки возвращаются в сомкнутое состояние [8, 9].

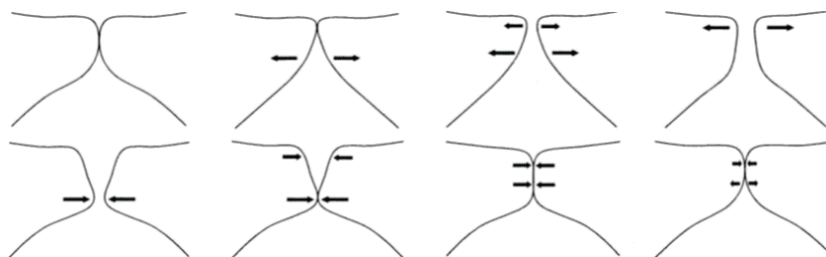


Рис. 2. Цикл колебания голосовых складок [10]

В зависимости от уровня активации мышц, при смыкании голосовых складок наблюдаются упругие деформации их тканей. Эти явления наглядно подтверждаются физическим моделированием процесса колебаний голосовых складок (рис. 3), выполненном при разных соотношениях модулей Юнга для внешнего и внутреннего слоёв складки [11]. Разнообразие упругих характеристик тканей голосовых складок связано с изменениями внутренней структуры, гипотонусе или гипертонусе мышц.

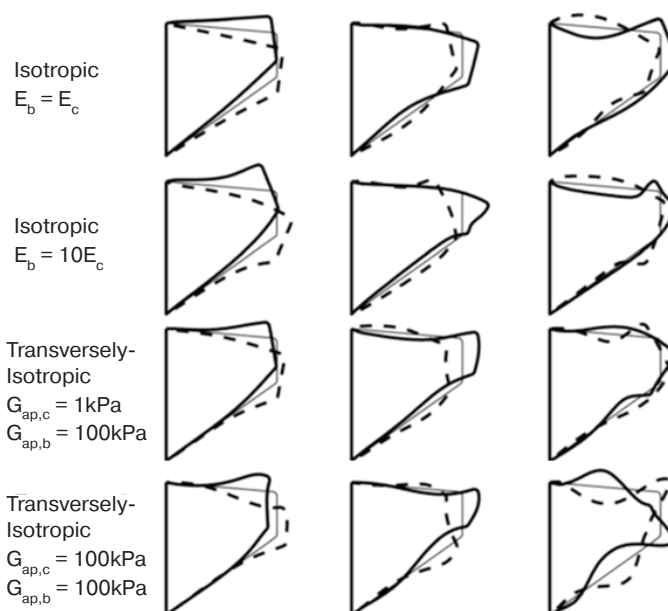


Рис. 3. Колебание голосовых складок при разных условиях [11]

Достаточно полное описание процессов колебаний голосовых складок достигается вычислительными средствами с использованием метода конечных элементов, что позволяет также учесть поверхностные волны. В [8] процесс колебаний во фронтальной плоскости идеализируется двумя собственными функциями, полученными как решение уравнений упругих деформаций. Пространственные моды этих колебаний показаны на рис. 4. Аналогичные собственные функции были получены в ходе физического моделирования процесса колебаний [12].

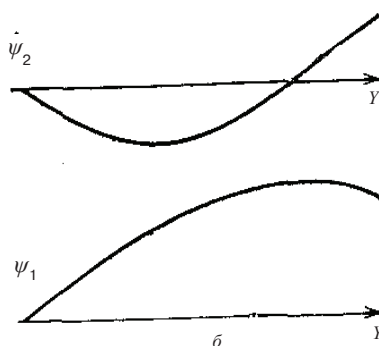


Рис. 4. Собственные функции колебаний голосовых складок во фронтальной плоскости [8]

#### Колебания голосовых складок в горизонтальной плоскости

При стробоскопии (рис. 5) наблюдаются горизонтальные колебания голосовых складок, обусловленные различием граничных условий на их концах [12–14].



Рис. 5. Колебания голосовых складок наблюдаемые при стробоскопии [16]

В модели [8] горизонтальные колебания описываются тремя основными собственными функциями (рис. 6), которые также наблюдаются при физическом моделировании процесса колебаний [12].

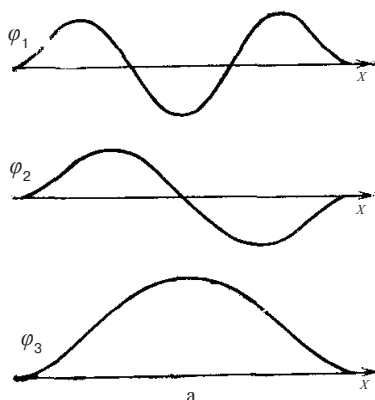


Рис. 6. Собственные функции колебаний свободных краев голосовых [8]

Физическое моделирование [12] показало, что вид собственной функции процесса горизонтальных колебаний зависит от жёсткости слоёв голосовой складки, скорости воздушного потока через голосовую щель и граничных условий (рис. 7). В ходе эксперимента, проведённого в [11], только одна модель (EPI, tension) позволила достигнуть скорости воздушного потока, лежащей в пределах 70–220 мл/с, что соответствует средней скорости потока, проходящего при фонации через голосовые складки человека [12].

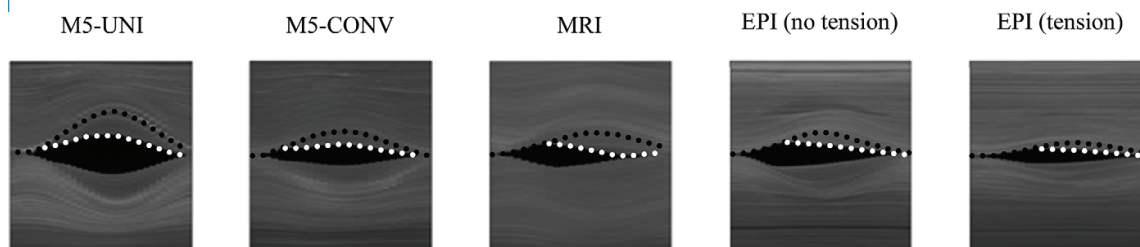


Рис. 7. Наблюдение колебаний свободных краев голосовых складок [12]

### Турбулизация воздушного потока

В результате расширения речевого тракта непосредственно над голосовой щелью воздушный поток завихряется на выходе из голосовой щели, что схематически показано на (рис. 8).

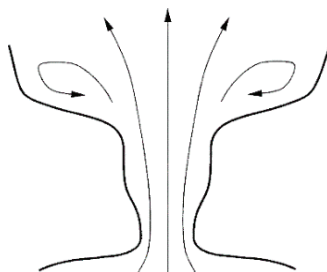


Рис. 8. Турбулентность воздушного потока

В [17] было проведено физическое моделирование для модели голосовых складок с динамическим управлением. В ходе эксперимента измерялась скорость прохождения воздушного потока через голосовую щель (рис. 9). Во время эксперимента наблюдалось отклонение воздушного потока в зависимости от остаточных вихревых сил и условий движения голосовых складок. Аналогичные результаты были получены при компьютерном 2D [18] [19] и 3D [20] моделировании.

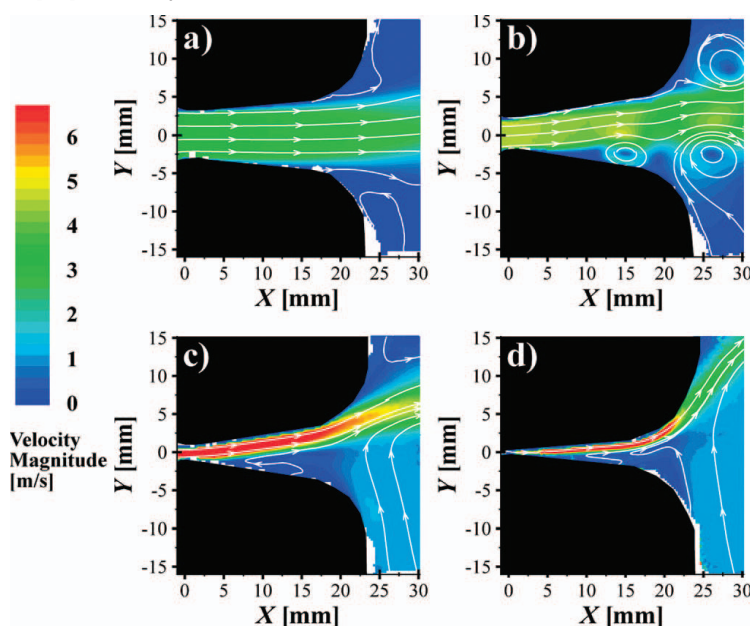


Рис. 9. Измерение скорости воздушного потока [17]

### Акустический анализ голоса

Известно несколько систем, применяемых в фониатрии для акустического анализа голоса и выявления патологий: Multi-Dimensional Voice Program (MDVP) [21], Praat [22] и Ling Waves [23]. В основу этих систем положен математический аппарат, позволяющий описать следующие эффекты:

- «Jitter» эффект — случайные колебания периода основного тона;
- «Shimmer» эффект — случайные колебания амплитуды сигнала на соседних периодах основного тона;
- турбулентный шум — уровень шума на периоде основного тона;
- нелинейность процесса колебаний голосовых складок.

Большинство параметров, описывающих эти эффекты, вычисляется на основе трека частоты основного тона (рис. 10).

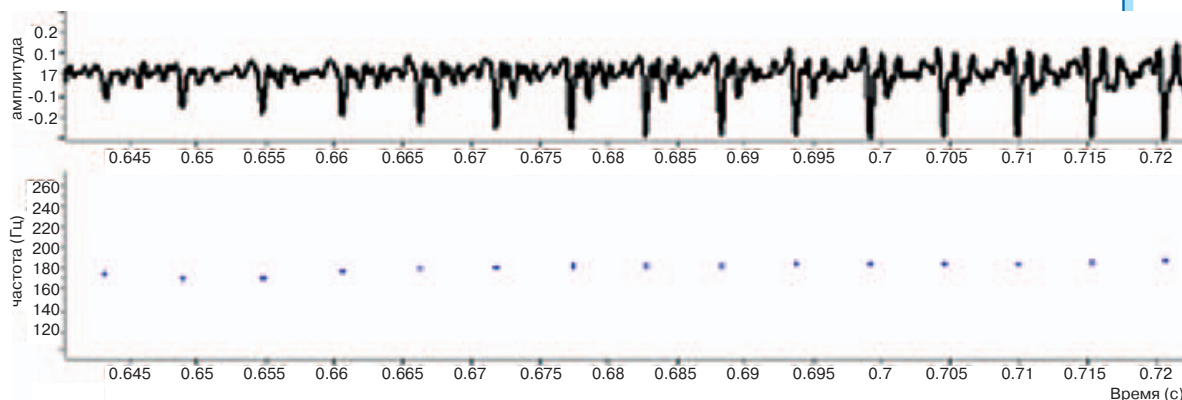


Рис. 10. Речевой сигнал (верхний график) и трек частоты основного тона (нижний график) сигнала

Обзоры различных подходов к определению частоты основного тона приводятся во множестве работ [24–27]. Все алгоритмы можно разделить на две группы:

- алгоритмы, основанные на анализе сигнала во временной области: ITU G.726 [28], YIN [29], TWIN [30];
- алгоритмы, основанные на анализе сигнала в спектральной области: REPS, DASH [31], TEMPO [32].

### «Jitter» эффект

Этот эффект заключается в случайных изменениях периода колебаний голосовых складок. Большинство характеризующих его параметров вычисляется на треке частоты основного тона (табл. 1).

Таблица 1

Параметры, характеризующие «Jitter» эффект

Параметр	Описание	Параметр	Описание
Mean F0 ( $F$ )	$\frac{1}{n} \sum_{i=1}^n F_i$	Jitter (%)	$\frac{MeanAbsoluteJitter}{F}$
Min F0	$\min(F_i)$	Max F0	$\max(F_i)$
Mean Absolute Jitter	$\frac{1}{n-1} \sum_{i=1}^{n-1}  F_{i+1} - F_i $	Pitch Perturbation Quotient	$\frac{100}{n-4} \sum_{i=3}^{i+2} \left  \frac{\sum_{k=i-2}^{i+2} F_k}{5} - F_i \right $
Standard derivation of F0 contour	$\frac{1}{n-1} \sum_{i=1}^n (F_i - \bar{F})^2$	Smoothed Pitch Perturbation Quotient	$\frac{100}{n-54} \sum_{i=28}^{i+27} \left  \frac{\sum_{k=i-27}^{i+27} F_k}{55} - F_i \right $
Phonatory Frequency Range	$\frac{12 \log \left( \frac{\max(F0)}{\min(F0)} \right)}{\log 2}$	Relative Average Perturbation	$\frac{100}{n-2} \sum_{i=2}^{i+1} \left  \frac{\sum_{k=i-1}^{i+1} F_k}{3} - F_i \right $
Pitch perturbation Factor	$\frac{100N_{F0 \geq threshold}}{N_{voice}}$	Directional Perturbation Factor	$\frac{100N_{\Delta \pm}}{N_{voice}}$

Кроме временного представления, в некоторых алгоритмах для вычисления «Jitter» эффекта используется спектральное представление сигнала, где оценивают отношение между гармонической и субгармонической компонентами спектра сигнала [33].

### «Shimmer» эффект

Этот эффект проявляется в случайных колебаниях амплитуды сигнала на соседних периодах основного тона. Большинство характеризующих его параметров также вычисляются на треке частоты основного тона (табл. 2). В норме амплитуда речевого сигнала меняется монотонно, тогда, как при патологии возникают резкие выбросы амплитуд [34].

Таблица 2

Параметры, характеризующие «Shimmer» эффект

Параметр	Описание	Параметр	Описание
Mean Amplitude	$\frac{1}{n} \sum_{i=1}^n A_i$	Max Amplitude	$\max(A_i)$
Min Amplitude	$\min(A_i)$	Shimmer (dB)	$\frac{1}{n-1} \sum_{i=1}^{n-1} 20 \log \left( \frac{A_i}{A_{i+1}} \right)$
Shimmer (%)	$\frac{MeanAbs.Shimmer}{\bar{A}}$	Amplitude Relative Average Perturbation	$\frac{100}{n-2} \sum_{i=2}^{n-1} \left  \frac{\sum_{k=i-1}^{i+1} A_k}{3} - A_i \right $
Mean absolute shimmer	$\frac{1}{n-1} \sum_{i=n-1}^l  A_{i+1} - A_i $	Amplitude Perturbation Quotient	$\frac{100}{n-4} \sum_{i=3}^{n-2} \left  \frac{\sum_{k=i-2}^{i+2} A_k}{5} - A_i \right $
Standard derivation of Amp contour	$\frac{1}{n-1} \sum_{i=1}^n (A_i - \bar{A})^2$	Smoothed Amplitude Perturbation Quotient	$\frac{100}{n-54} \sum_{i=28}^{n-27} \left  \frac{\sum_{k=i-27}^{i+27} A_k}{55} - A_i \right $
Amplitude Perturbation Factor	$\frac{100N_{A>threshold}}{N_{voice}}$	Amplitude Directional Perturbation Factor	$\frac{100N_{\Delta\pm}}{N_{voice}}$

### Турбулентный шум

Турбулентный шум — неотъемлемая компонента речевого сигнала. Уровень шума определяет качество голоса в целом и характеризует наличие хрипов и осиплости голоса. Используют следующие параметры, позволяющие характеризовать уровень турбулентного шума:

— Turbulent noise index [35]:

$$TNI = 100 \left( 1 - \frac{1}{N-1} \sum_{n=1}^{N-1} R(t_n, T_n) \right),$$

где  $N$  — длина трека основного тона,  $R(t_n, T_n)$  — нормализованная автокорреляционная функция;

— Soft phonation index [36] — определяется отношением гармонической энергии в частотном диапазоне от 70Hz до 1600Hz к гармонической энергии в частотном диапазоне от 1600Hz до 4500Hz [21];

— Voice turbulence index — характеризуется средним отношением негармонической энергии сигнала в частотном диапазоне от 2800Hz до 5800Hz к гармонической энергии в частотном диапазоне от 70Hz до 4500Hz. При этом уровень гармонической энергии вычисляется в области с минимальным уровнем колебания частоты гармоник, т.е. амплитуды сигнала и минимальной энергией субгармонической составляющей сигнала [21] [36];

— Harmonic to Noise Ratio — характеризует относительный уровень шума в речевом сигнале. Способы математического определения данного параметра различны в разных исследованиях [37 — 43]. Наиболее распространены следующие методы вычисления:

$$HNR(dB) = 10 \log \frac{r'_x(t_{\max})}{1 - r'_x(t_{\max})},$$

где  $r'_x(t)$  — автокорреляционная функция,  $t_{\max}$  — момент времени, на котором  $r'_x(t)$  принимает максимальное значение;

$$HNR_{\text{yum}} = \frac{n \int_0^{T_i} f_A^2(t) dt}{\sum_{i=0}^n \int_0^{T_i} (f_i(t) - f_A(t))^2 dt},$$

где  $f_i(t)$  — сигнал на  $i$ -ом периоде основного тона,  $f_A(t)$  — сигнал, равный среднему от всех сигналов на периоде основного тона;

$$HNR = \frac{K |HNR_{\text{yum}} (N-1) - 1|}{N(1 - HNR_{\text{yum}} (K-1))}; K = \frac{N \sum_{i=1}^N A_i^2}{(\sum_{i=1}^N A_i)^2},$$

где  $N$  — количество определённых периодов основного тона,  $A_i$  — максимальная амплитуда сигнала на  $i$ -ом периоде основного тона.

В программной системе MDVP этот параметр вычисляется как отношение негармонической спектральной энергии на интервале от 1500 Hz до 4500 Hz к гармонической энергии на интервале от 70 Hz до 4500 Hz [21];

• Glottal to Noise Excitation Ratio — характеризует качество речевого сигнала с учётом влияния формантных частот. Сначала с помощью обратной фильтрации из речевого сигнала удаляются спектральные компоненты в области полюсов передаточной функции речевого тракта. Далее рассчитывается кросскорреляционная функция для гильбертовых огибающих в различных частотных диапазонах (рис. 11).

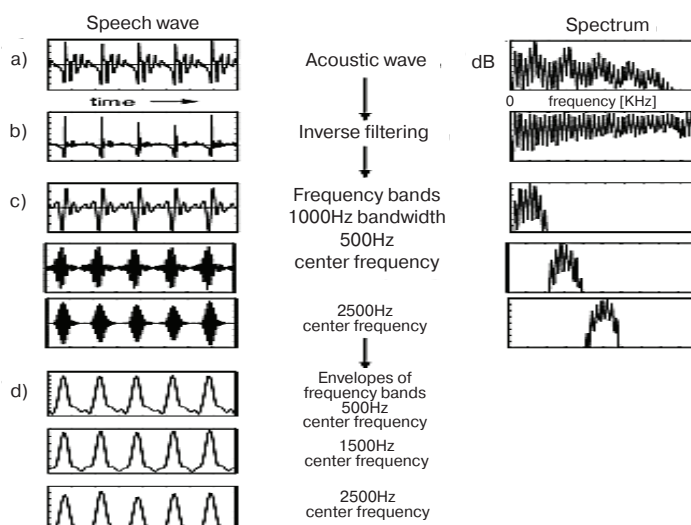


Рис. 11. Схема алгоритма GNER [44]



Этот параметр позволяет определить уровень турбулентного шума на определённых участках речевого сигнала [44]. Важной особенностью алгоритма является отсутствие необходимости в расчёте частоты основного тона [45].

### Нелинейный анализ

Процесс голосообразования — нелинейный процесс [46]. На колебания влияют аэродинамические, биомеханические, физиологические и акустические факторы. Нелинейность колебаний приводит к смещению фаз и нестационарности колебаний голосовых складок [47] [48] [49] [46] [50]. Наличие патологии значительно усиливает нелинейность колебаний. Свойства речевого сигнала как динамической системы наглядно демонстрируются на фазовой плоскости [51] (рис. 12).

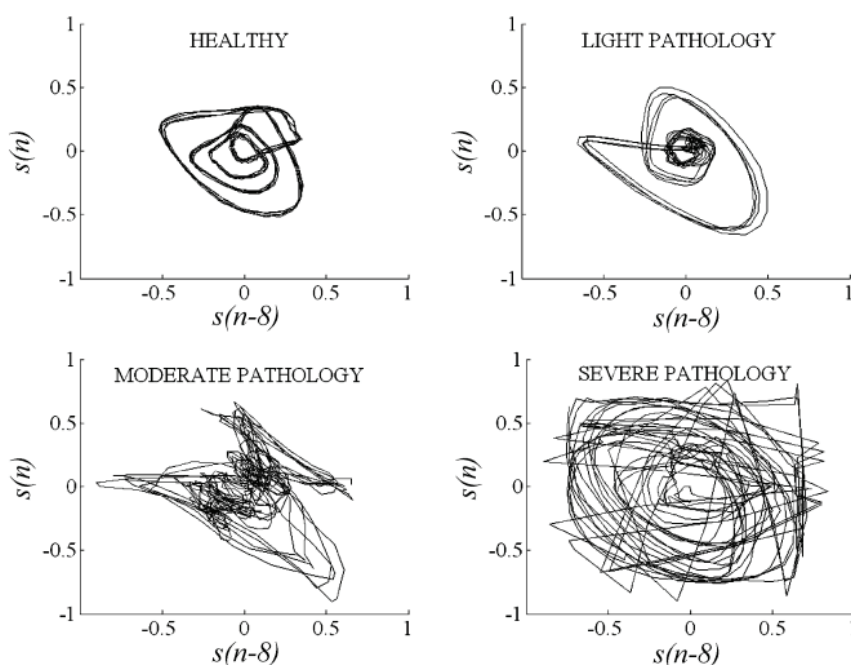


Рис. 12. Аттракторы различных типов голоса для гласного /а/ продолжительностью 30-ms [51]

Для описания нелинейности системы используются такие методы, как энтропия Шеннона [52] [53], энтропия Ренья [54], FMMI [55], метод корреляционной размерности («Taken's Estimator») [56] и т.д.

### Оценка физиологических характеристик состояния голосовых складок

В результате нарушения процесса колебаний голосовых складок возникают «Jitter», «Shimmer» эффекты, усиливается влияние турбулентного шума и нелинейность процесса колебаний в целом. Проявление подобных эффектов позволяет предположить наличие патологии голосовых складок. При этом все параметры, применяемые для описания эффектов, лишь косвенно характеризуют процесс колебаний и не дают никакой информации о физических свойствах голосовых складок. Использование физически адекватных параметров не только увеличивает вероятность правильной классификации «норма/патология», но и значительно упрощает интер-

претацию результатов при обследовании пациентов. Врач может описывать состояние складок, а не процессы, возникающие в результате их колебания.

В работе [57] был продемонстрирован метод получения функции площади голосовой щели, на основе которого восстанавливались параметры одномассовой модели голосовых складок, позволяющей описать эффект их инерционности [57]. Предполагается, что голосовая складка имеет сосредоточенную массу и совершает только горизонтальные колебания (рис. 13).

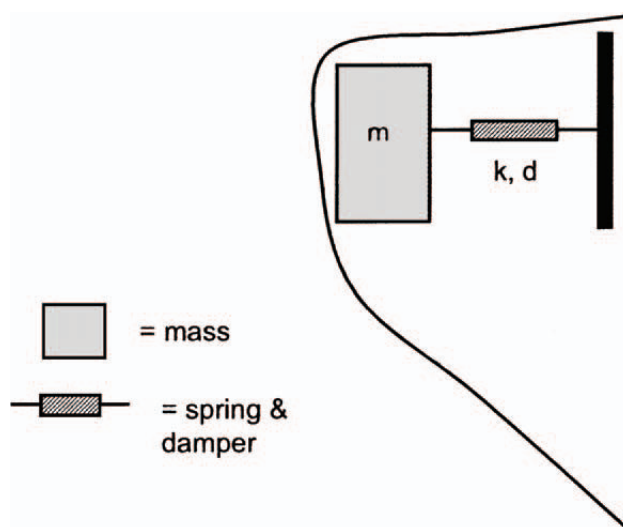


Рис. 13. Одномассовая модель голосовой складки

Модель представляет собой систему «масса-пружина-демпфер», которая совершает колебания под воздействием аэродинамической силы, возникающей в результате увеличения подвязочного давления. Динамика системы описывается следующим уравнением:

$$F(t) = m\ddot{x} + d\dot{x} + kx,$$

где  $m$  — масса складки,  $d$  — постоянная характеристика пружины,  $k$  — постоянная характеристика глушителя,  $x(t)$  — смещение массы  $m$ .

### Результаты применения методов акустического анализа

Для определения ошибки при классификации «норма/патология» необходимо иметь обширную базу данных, включающую в себя запись голосов в норме и патологии. Такая база была собрана на базе Massachusetts Eye and Ear Infirmary (MEEI) Voice and Speech Lab. Она включает в себя 1 500 записей голосов в норме и различной степени патологии. В большинстве случаев для тестирования используют усечённый вариант этой базы данных: 53 записи для нормы и 173 записи для патологии. Некоторые исследования проводятся на собственных базах данных, что значительно затрудняет интерпретацию и сравнение их результатов.

Результаты различных исследований (таблица 3) показывают ошибку классификации «норма/патология» от 20% до 2%. Как правило, на них влияет не только информативность исследуемого параметра, но и алгоритм, применяемый для этой классификации. Дополнительные разногласия в оценке методов анализа голоса частично вызваны влиянием пола и возраста диктора, темпа речи, типа артикуляции и привычной для диктора частоты основного тона на объективные акустические параметры речевого сигнала.

Таблица 3

## Результаты классификации «норма/патология» [51]

Автор	База данных Норма + патология	Измеряемые параметры	Классификатор	Вероятность правильной классификации
Linder [58]	8 + 112	Jitter, Shimmer, Standard deviation of fundamental frequency, Glottal-to-noise excitation ratio	Neural Network	80%
Wallen [59]	9 + 20	Perturbation, cepstral, LPC	Multi-layer perceptron	85,8%
Boyanov [60]	50 + 150	Perturbation, noise, energies	K-nearest neighbors, linear discriminant analysis, self-organized map	93,5%
Alonso [61]	100 + 68	Noise, periodicity, chaos, shimmer, spectral	Neural Network	92,76%
Parsa [62]	MEEI 53 + 173	Noise	Linear discriminant	98,7%
Saenz-Lechon [63]	MEEI 53 + 173	Mel-Frequency cepstral Coefficient	Multi-layer perceptron	89,6%
Hadjitodorov [64]	MEEI 53 + 638	Perturbation, noise	K-nearest neighbors	96,1%
Henriquez [50]	MEEI 53 + 173	Chaos	Neural Network	82,47%
Pedro Gómez-Vilda [56]	100 + 100	Glottal Source signature	Gauss mixture model	92%

**Обсуждение и заключение**

Как видно из приведённой таблицы 3, при определённых условиях достигается весьма высокая вероятность правильной классификации фактора «норма/патология». Вместе с тем, эти результаты получены для достаточно выраженных стадий заболевания. Большинство исследований производились на различных базах данных, имеющих малую выборку из голосов в норме и патологии, что значительно усложняет сравнение результатов и ставит под сомнение устойчивость таких высоких оценок в реальных условиях эксплуатации.

Постановка задачи максимально ранней диагностики отклонения от нормы требует разработки новых методов анализа речевого сигнала.

Большинство параметров, используемых при описании текущего состояния голосовых складок, лишь косвенно описывают процесс колебаний и не дают никакой информации о физических свойствах голосовых складок, что затрудняет интерпретацию результатов врачами-специалистами. Совместное использование косвенных и физиологических параметров может значительно улучшить результаты классификации и стабильность работы систем в реальных условиях эксплуатации.

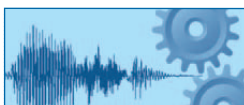
При получении физически адекватных параметров особое внимание следует уделить выбору модели колебаний голосовых складок. Решением данной проблемы может являться замена непрерывной модели, описанной, например в [65], на параметрическую. При этом сохраняется физиологическая адекватность модели голосового источника, значительно уменьшаются требования к вычислительным ресурсам, и увеличивается число параметров, характеризующих состояние голосовых складок.

Задача вычисления параметров голосового источника является некорректной обратной задачей, т.е. задачей, формально не имеющей ни единственного,

ни устойчивого решения [66]. Для приближённого решения такой задачи необходимо располагать дополнительной информацией. Так, помимо выбора физиологически адекватной модели голосового источника, необходимо по экспериментальным данным установить ограничения на параметры модели и найти критерии оптимальности решения обратной задачи.

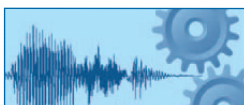
## Список литературы

1. International Agency for Research on Cancer, «Cancer Worldwide,» World Health Organization, 2011.
2. Крехно О.П., Фролов А.С., III конгресс с международным участием «Опухоли головы и шеи». Реабилитация голосовой функции после ларингэктомии по поводу рака гортани с помощью протезирования. Сочи, 2009.
3. Becker W., Naumann H.H., Faltz C.R. Ear, Nose and Throat Diseases. Thieme Medical Publishers, 1994.
4. Hetrich I., Lutzenberger W., Spieker S., Ackhermann H. Fractal dimension of sustained vowel productions in neurological disphonias: An acoustic and electroglottographic analysis // JASA, 1997. V. 102. № 1. P. 652–654.
5. Klingholtz F. Acoustic recognition of voice disorders: A comparative study of running versus sustained vowels // JASA, 1990. V. 85. № 5. P. 2218–2224.
6. Gray S.D. Cellular physiology of the vocal folds // Otolaryngol Clin North Am., 2000. V. 33. № 4. P. 679–698.
7. Нейман Л.В., Богомильский М.Р. Анатомия, физиология органов слуха и речи: Учебное пособие для студентов педагогических учебных заведений. Москва: Владос, 2001.
8. Сорокин В.Н. Теория Речеобразования. Москва: Радио и Связь, 1985.
9. Titze I R. Principles of Voice Productions. Prentice-Hall, 1994.
10. Zhang Z. Restraining mechanisms in regulating glottal closure during phonation // JASA, 2011. V. 130. № 6. P. 4010–4019.
11. Murray P.R., Thomson S.L. Vibratory responses of synthetic, self-oscillating vocal fold models // JASA, 2012. V. 132. № 5. P. 3428–3438.
12. Shipp T. Aspects of voice production and motor control. *Speech Motor Control*, 1982. P. 105–112.
13. Sonesson B. Vocal fold kinesiology. *Speech Motor Control*, 1982. P. 113–118.
14. S. T. Vocal fold physiology conference. The function of the extrinsic laryngeal muscles, Kurume, 1980.
15. Baken R.J., Orlikoff R.F. Clinical Measurement of Speech and Voice 2nd ed, San Diego: Singular, 2000.
16. Erath B.D., Peterson S.D., Zanartu M., Wodicka G.R., Plesniak M.W. A theoretical model of the pressure field arising from asymmetric intraglottal flows applied to a two-mass model of the vocal folds // JASA, 2011. V. 130. № 1. P. 389–403.
17. Xue Q., Mittal R., Zheng X., Bielamowicz S. Computational modeling of phonatory dynamics in a tubular three-dimensional model of the human larynx // JASA, 2012. V. 32. № 3. P. 1602–1613.
18. Zheng X., Mittal R., Bielamowicz S., Luo H. A computational study of the effect of false vocal folds on glottal flow and vocal folds vibration during phonation // Ann. Biomed. Eng., 2009. V. 37. № 3. P. 625–642.



19. Zheng X., Mittal R., Xue Q., Bielałowicz S. Direct-numerical simulation of the glottal jet and vocal-fold dynamics in a three-dimensional laryngeal model // *JASA*, 2011. V. 130. № 1. P. 404–415.
20. An Xue S. Effects of aging on selected acoustic voice parameters: preliminary normative data and educational implications. 2001.
21. Boersma P. Optimality-theoretic learning in the Praat program. IFA Proceedings, 1999. V. 23. P. 17–35.
22. «LingWAVES», WEVOSYS: [сайт] URL: <http://www.wevosys.com/products/lingwaves/lingwaves.html>
23. Rabiner L.R., Cheng M.J., Rosenberg A.E., McGonegal C.A. A comparative performance study of several pitch detection algorithms. *IEEE Trans. Audio Electroacoust.*, 1976. P. 399–417.
24. Gerhard D. Pitch extraction and fundamental frequency: history and current techniques. University of Regina, Saskatchewan, Canada, 2003.
25. De Cheveigne A. International Conference on Acoustics. Pitch perception models from origins to today, Kyoto, 2004.
26. Sorokin V.N., Trifonenkov V.P. Autocorrelational Analysis of Speech Signal, 1996. V. 3. № 42.
27. ITU Recommendation G.726: [сайт] URL: <http://www.itu.int/rec/T-REC-G.726/en>
28. De Cheveigne A, Kawahara H. YIN, a fundamental frequency estimator for speech and music // *JASA*, 2002. № 111. P. 1917–1930.
29. Цыплихин А.И. Анализ импульсов голосового источника // *Акустический журнал*, 2007, Т. 53. С. 119–133.
30. Nakatani T., Irino T. Robust and accurate fundamental frequency estimation based on dominant harmonic components // *JASA*, 2004. V. 116. № 6. P. 3690–3700.
31. Kawahara H., Masuda-Kasuse I., De Cheveigne A. Restructuring speech representations a pitch-adaptive time-frequency smoothing and instantaneous-frequency-based F0 extraction: Possible role of repetitive structure in sounds // *Speech Communication*, 1999. V. 29. № 3–4. P. 187–207.
32. Vasilakis M., Stylianou Y. Voice Pathology Detection Based on Short-Term Jitter Estimations in Running Speech // *Folia Phoniatrica et Logopaedica*, 2009. V. 1. № 509.
33. Kitajima K. Vocal Shimmer in Sustained Phonation Normal and Pathology Voice // *Studia phonologica*, 1976. V. 10. P. 23–27.
34. Mitev P., Hadjitodorov S. A method for turbulent noise estimation in voiced signal // *Med Biol Eng Comput.*, 2000. V. 38. № 6. P. 625–631.
35. Nicola V.D., Fiorella M.I., Spinelli D.A., Fiorella R. Acoustic analysis of voice in patients treated by reconstructive subtotal laryngectomy. Evaluation and critical review // *ACTA OTORHINOLARYNGOL*, 2006. V. 26. P. 56–68.
36. Oller L.L. Analysis of voice signals for the harmonic-to-noise crossover frequency. UPC Barcelona, 2008.
37. Shama K., Krishna A., Cholayya N.U. Study of harmonics to noise ratio and critical-band energy spectrum of speech as acoustic indicators of laryngeal and voice pathology. *EURASIP Journal on Advances Signal Processing*, 2007.

38. *Yingyong Q.* Temporal and spectral estimations of harmonic to noise ratio in human voice signal // *JASA*, 1997. V. 102. № 1. P. 537–543.
39. *Girin. L.* 14th European Signal Processing Conference (EUSIPCO 2006) // Theoretical and experimental bases of a new method for accurate separation of harmonic and noise components of speech signals, Florence, Italy, 2006.
40. *Boersma P.* Accurate short-term analysis of the fundamental frequency and the harmonic to noise ratio of sampled sound // *Proceedings*, 1993. V. 17. P. 97–110.
41. *Yumoto E., Gould W. J.* Harmonics to noise ratio as an index of the degree of hoarseness // *JASA*, 1982. V. 71. № 6. P. 1544–1550.
42. *Ferrer C., Gonzalez E., Hernandez-Diaz H., Torres D., del Toro A.* Removing the influence of shimmer in the calculation of harmonic noise ratios using ensemble-averages in voice signals // *EURASIP Journal on Advances in Signal Processing*, 2009.
43. *Michaelis D., Gramss T., Strube H.* Glottal to noise excitation ratio — a new measure for describing pathological voices // *Acta acustica*, 1996. V. 83. P. 700–706.
44. *Frochlich M., Michaelis D., Strube H.W., Kruse E.* Advances in Quantitative Laryngoscopy, 2nd 'Round Table' // *Acoustic voice quality description: Case studies for different regions of the hoarseness diagram*, Erlangen, 1997.
45. *Herzel H., Berry D., Titze I.R., Saleh M.* Analysis of vocal disorders with methods from nonlinear dynamics // *Journal of Speech & Hearing research*, 1994. V. 37. P. 1008–1019.
46. *Tokuda I.P., Horacek J., Svec J.G., Herzel H.* Bifurcations and chaos in register transitions of excised larynx experiments // *Chaos*, 2008. № 18.
47. *Zhang Y., McGilligan C., Zhou L., Vig M., Jiang J.J.* Nonlinear dynamic analysis of voice before and after surgical excision of vocal polyps // *JASA*, 2004. V. 115. № 5. P. 2270–2277.
48. *Jiang J.J., Zhang Y.* Chaotic vibration induced by turbulent noise in a two-mass model of vocal folds. *Acoustical Society of America*, 2002.
49. *Herzel H., Švec J.G., Horáček J., Tokuda I.T.* Bifurcations and chaos in register transitions of excised larynx experiments // *Chaos*, 2008. № 18.
50. *Henriquez P., Alonso J.B., Ferrer M.A., Travieso C.M., Godino-Llorente J.I., Diaz-de-Maria F.* Characterization of Healthy and Pathological Voice Through Measures Based on Nonlinear Dynamics // *IEEE transactions on audio, speech, and language processing*, 2009. V. 17. № 6. P. 1186–1195.
51. *Shannon C.E.* A mathematical theory of communication // *Bell System Technical Journal*, 1948. V. 27. P. 379–423.
52. *May E.C., James S., James C.L.* Toward a Reductionist Model of Anomalous Cognition // *Shannon Entropy as an Intrinsic Target Property*, 1994. V. 2. № 22.
53. *Renyi A.* *Proceedings of the 4th Berkeley Symposium on Mathematics* // On measures of information and entropy, Berkeley, 1960.
54. *Fraser A.M., Swinney H.L.* Independent coordinates for strange attractors from mutual information // *Phys. Rev. A*, 1986. V. 33. P. 1134.
55. *Kantz H., Schreiber T.* *Nonlinear time series analysis*, Cambridge: Cambridge Univ. Press., 1997.
56. *Gomez-Vilda P., Fernandez-Baillo R., Rodllar-Biarge V., Lluís V.N., Alvarez-Marquina A., Mazaira-Fernandez L.M., Martinez-Olalla R., Godino-Llorente J.I.* Glottal source biomedical signature for voice pathology detection. *Speech Communication*, 2008.
57. *Flanagan J.L., Landgraf L.* Self-oscillating source for vocal tract synthesizers // *IEEE Trans. Audio Electroacoust.*, 1968. V. 16. P. 57–64.



58. Linder R., Albers A.E., Hess M., Poppl S.J., Schonweiler R. Artificial neural network-based classification to screen for dysphonia using psychoacoustic scaling of acoustic voice features // *J Voice*, 2008. V. 22. P. 155–163.
59. Wallen E.J., Hansen J.H.L. A screening test for speech pathology assessment using objective quality measures // *Proc. Int. Conf. Spoken Lang. Process.*, 1996. V. 2. P. 776–779.
60. Boyanov B., Hadjitodorov S. Acoustic analysis of pathological voices: A voice analysis system for screening of laryngeal diseases // *IEEE Eng. Med. Biol. Mag.*, 1997. V. 16. P. 74–82.
61. Alonso J.B., Díaz-de-María F., Travieso C.M., Ferrer M.A. Using nonlinear features for voice disorder detection // *Proc. 3rd Int. Conf. Nonlinear Speech Process.*, 2005. V. 1. P. 94–106.
62. Parsa V., Jamieson F.G. Identification of pathological voices using glottal noise measures // *Journal of Speech, Language and Hearing Research*, 2000. V. 30. P. 529–538.
63. Saénz-Lechón N., Godino-LLorente J.I., Osma-Ruiz V., Gómez-Vilda P. Methodological issues in the development of automatic systems for voice pathology detection // *Biomed. Signal Process. Control.*, 2006. V. 1. № 2. P. 120–128.
64. Hadjitodorov S., Mitev P. A computer system for acoustic analysis of pathological voices and laryngeal diseases screening // *Med. Eng. Phys.*, 2002. V. 24. P. 419–429.
65. Titze I.R. *The Myoelastic Aerodynamic Theory of Phonation*, National Center for Voice and Speech, 2006.
66. Сорокин В.Н. *Речевые процессы*, М.: Народное образование, 2012.

#### Сведения об авторе:

Лысак А.П. —

аспирант Института проблем передачи информации РАН.