

О. П. Скляр

НИИ уха, горла, носа и речи

Россия, 190013, Санкт-Петербург, ул. Бронницкая, 9, e-mail: skljarov@admiral.ru

Фракталы и крупномасштабная временная структура акустического речевого сигнала и музыки

Получена 30.10.2004, опубликована 04.12.2004

В статье рассматриваются Хаусдорфова размерность и Колмогоровская энтропия последовательности длительностей голосовых сегментов в речи (V-ритма). Хаусдорфовы размерности и Колмогоровская энтропия вычисляются из экспериментально определенных V-ритмов образцов как нормальной речи, так и речи при заикании. Эти вычисления оказались возможными благодаря введению в рассмотрение обобщенной размерности Реньи и обобщенной энтропии Реньи. В статье показано, что V-ритм речи образует фрактальную структуру, причем соответствующая Хаусдорфова размерность больше топологической размерности. В свою очередь, фрактальная размерность нормальной речи больше фрактальной размерности участков речи, состоящих из запинок. Кроме того, обнаружено, что Колмогоровская энтропия ритма нормальной речи положительна и конечна, что является необходимым и достаточным условием существования детерминированного хаоса в структуре ритма. Нижняя граница Колмогоровской энтропии для запинок отрицательна, что говорит в пользу регулярности структуры запинок в речи. На примере пьес Вивальди и Паганини показано, что Хаусдорфова размерности музыки ниже размерности запинок, что позволяет сравнить структуру V-ритма музыки со структурой ритма ранней фонетики у детей.

1. ВВЕДЕНИЕ

В последнее десятилетие значительное число задач в нелинейной акустике было успешно решено методами теории критических явлений [1]. Довольно часто динамика таких задач вблизи критической точки описывается средствами теории динамического хаоса.

Оказалось, что этот подход является весьма продуктивным при объяснении многих эффектов при описании крупномасштабной временной структуры акустических речевых сигналов [2, 3]. Можно показать, исходя из исследования структуры сигнала, что речь обладает целым набором динамических режимов. Переход из одного режима в другой может происходить под влиянием как внешних, так и внутренних воздействий на динамическую систему. В нашем случае такой системой является механизм речеобразования человека.

В работе мы описываем метод регистрации крупномасштабной временной структуры речевого сигнала, базирующийся на сегментации речевого сигнала по принципу «есть голос — нет голоса». Пронумерованную последовательность длительностей сегментов в речи, когда происходят колебания голосовых складок, мы назвали V-ритмом. На временной оси такой ритм образует некий аналог канторова множества. Примем длину всего образца речи за единицу. Тогда, как можно допустить на основе экспериментальных данных, обычная топологическая размерность такого множества V-сегментов имеет размерность строго меньше 0,7 [6]. Однако, как будет показано в статье, нижние границы хаусдорфовых размерностей V-ритмов, оцененные по экспериментальным последовательностям V-сегментов на обширной фонотеке сигналов, оказались в интервале (0,7...0,9).

Мандельброт предложил называть фракталами множества, для которых Хаусдорфова размерность строго больше топологической [4]. Таким образом, в статье показывается, что V-ритм речи как в норме, так и при заикании, фрактален.

Измерялась также безмодельным способом Колмогоровская энтропия таких множеств. На основе этих вычислений мы заключаем, что система может иметь как регулярные, квазипериодические свойства, характерные для участков речи с нарушенным ритмом, так и иррегулярные хаотические свойства, характерные для нормальной речи. Участки речи с нарушенным ритмом можно получить, регистрируя речь заикающихся.

Оказалось, что система речеобразования имеет критическую точку в сценарии перехода динамики V-ритма к хаосу в зависимости от внешнего управляющего параметра. В качестве такого сценария, как оказалось, можно принять сценарий Фейгенбаума перехода к хаосу в зависимости от управляющего параметра. Управляющий параметр может использоваться для того, чтобы приводить такие временные характеристики, как речевой V-ритм, в область нормальных значений.

2. СЕГМЕНТАЦИЯ РЕЧЕВОГО СИГНАЛА ПО ПРИНЦИПУ “VOICE/UNVOICE”

Производилась регистрация акустических речевых сигналов, полученных как от нормальных в речевом отношении лиц, так и от заикающихся. Исследования проводились в течение двух лет. Речь регистрировалась как при чтении стандартного текста объемом в 120 слогов, так и при произнесении спонтанной речи длительностью около 1,5 минут в группе заикающихся, в которую входили как взрослые мужчины и женщины, так и дети от 4 лет и старше (у детей дошкольного возраста регистрировалась только спонтанная речь). Число обработанных записей, производимых трижды в течение месячного курса коррекции, в течение двух лет составило 600 образцов. Число контрольных записей речевого сигнала составило 100 образцов как взрослой, так и детской речи без явных признаков патологии.

Речевые сигналы вводились в компьютер с частотой дискретизации 10000 Гц. После предварительной фильтрации определялся порог сегментации для каждого отдельного субъекта с помощью тестовой фразы. Амплитудный порог нарастал постепенно с

достаточно малым шагом до тех пор, пока впервые не выделялись шесть (и только шесть) сегментов, содержащихся в тестовой фразе «папа, папа, папа». Для определения момента, когда начинается не голосовой сегмент, использовался параметр временного разрешения, составляющий 40 мсек. Процесс определения порога прекращался, если коэффициент вариаций длительностей сегментов не превышал 0,5.

Затем это значение порога использовалось для автоматической сегментации основного сигнала. Сравнение результатов, полученных с помощью автоматической сегментации, и результатов как ручной сегментации сигнала, так и результатов, полученных при вычислении интонационного контура, дали удовлетворительное согласие. Кроме того, контрольное прослушивание сигнала, полученного обнулением «не голосовых» сегментов сигнала, не приводило к нарушению разборчивости речи.

Среднюю длительность этих сегментов мы назвали индикатором «V/U»-темпа, а их последовательность — «V/U»-ритмом. В случае только «Voice» сегментов эти величины называются соответственно индикатором «V»-темпа и «V»-ритмом.

3. ФРАКТАЛЬНО-ХАОТИЧЕСКАЯ ПРИРОДА РЕЧИ

3.1. Хаусдорфовы размерности

Реконструкция динамики сложных систем возможна безмодельным способом по информации, содержащейся во временной экспериментальной последовательности только одной, но существенной переменной. В нашем случае такой последовательностью являются элементы «V/U»-ритма или «V»-ритма. Ренормгрупповые свойства фракталов обеспечивали равноправность этих двух представлений ритма [4].

В частности, для «V»-ритма возможно реконструировать хаусдорфовы размерности исследуемого множества. В непрерывном случае эта размерность соответствует размерности странного аттрактора [5, стр. 109–110], к которому стягивается фазовая траектория динамической задачи.

Следуя Шустеру [5], проверим, является ли V-ритм речи образованием с дробной Хаусдорфовой размерностью. Для этого сопоставим V-ритму $\{T_1, T_2, \dots, T_n\}$ аналог канторовского множества единичной длины. Полученные нами экспериментальные данные позволяют предположить, что топологическая размерность такого аналога как для заикающихся, так и в случае нормальной речи, строго меньше 0,7 [6].

В соответствии с [5], приведем полученное множество длины L (в миллисекундах) к единичной длине и будем проверять это множество в контрольных точках через интервалы времени в $1/L$. Разделим одномерное пространство на ячейки $l=1000/L$. Размер ячеек оказался достаточным для достоверного определения Хаусдорфовой размерности по алгоритму Grassberga и Procaccia [7].

Хаусдорфова размерность (вернее, ее нижняя граница), вычисленная напрямую по этому алгоритму, как оказалось, варьирует на огромном множестве сигналов на

интервале (0,7...0,9). Большая статистика экспериментов свидетельствует в пользу надежности результатов.

В соответствии с алгоритмом Grassberga и Procaccia, для определения хаусдорфовой размерности D некоторого множества, занимающего область с объемом L^D в D -мерном пространстве, покроем это множество кубами с объемом l^D . Минимальное число таких непустых кубов, покрывающих множество, есть $M(l) = L^D (1/l)^D$. Из этого выражения можно получить приближенную оценку D :

$$D = \lim_{l \rightarrow 0} \left[\frac{\ln M(l)}{\ln(1/l)} \right]. \tag{1}$$

На практике более удобно для оценки этой размерности использовать математическую конструкцию, известную как размерность Реньи, D_f , связанную с вероятностью p_i нахождения контрольной точки в i -ой ячейке в степени f :

$$D_f = \lim_{l \rightarrow 0} \left(\frac{1}{f-1} \right) \frac{\ln \left(\sum_{i=1}^{M(l)} p_i^f \right)}{\ln l}, \quad f = 0, 1, 2, \dots \tag{2}$$

При $f \rightarrow 0$ мы имеем из формулы (2):

$$D_0 = - \frac{\lim_{l \rightarrow 0} \left(\ln \sum_{i=1}^{M(l)} 1 \right)}{\ln l} = - \lim_{l \rightarrow 0} \frac{\ln M(l)}{\ln l} = D, \tag{3}$$

т. е. размерность Реньи D_0 совпадает с хаусдорфовой размерностью (1). В силу монотонности D_f как функции f размерность Рени уменьшается как функция степени и, поэтому, выполняется следующее неравенство: $D_2 \leq D_0 = D$. Таким образом, наибольшая нижняя граница Хаусдорфовой размерности D_2 представима в виде

$$D_2 = \lim_{l \rightarrow 0} \frac{\ln \left(\sum_{i=1}^{M(l)} p_i^2 \right)}{\ln l}. \tag{4}$$

Принимая во внимание, что вероятность нахождения контрольной точки в i -ой ячейке p_i оценивается как $p_i = \lim_{N \rightarrow \infty} N_i / N$, где N есть общее число контрольных точек через интервалы $1/L$, и N_i есть число точек в i -ой ячейке, формула (4) может быть рассчитана из экспериментально измеренных длительностей сегментов. На практике наибольшую нижнюю границу размерности D_2 можно вычислить как тангенс угла наклона линейной регрессии следующих точек $\left(\ln \frac{1}{N^2} \left(\sum_i^{M(l)} N_i^2 \right); \ln(l) \right)$, вычисленных при разных l .

Типичные длины образцов речи L представляют собой величины порядка $(150...200)10^3$ миллисекунд, но могут быть и больше. В качестве наименьшей ячейки l_0 покрытия мы выбрали 10^3 миллисекунд, все последующие покрытия l были кратны этому l_0 . Вероятностные отношения N_i/N вычислялись как отношения числа миллисекунд V -сегментов, попавших в i -ую ячейку, к общему числу миллисекунд данного образца речи. Для определения фрактальной размерности была написана подпрограмма (на языке Паскаль) к программе сегментации речи, так что при каждой сегментации сигнала с целью определения ритма мы автоматически получали оценку его фрактальной размерности.

Эти хаусдорфовы размерности для нормальной речи и для речи с нарушенным ритмом оказались равными: нормальный ритм: $D_2=0,85 \pm \sigma_\Sigma=0,85 \pm 0,1$; нарушенный ритм: $D_2=0,75 \pm \sigma_\Sigma=0,75 \pm 0,03$. Дробная размерность больше нуля для нарушенного ритма $D_2=0,75$ есть доказательство его фрактальной природы. Можно показать, что раннему периоду становления фонетики у детей соответствуют еще меньшие значения фрактальных размерностей [8].

Хаусдорфова размерность для ритма нормальной речи соответствует хаотической природе ритма. Это соответствие доказывается безмодельным способом с помощью энтропии Колмогорова.

Нами была предпринята попытка вычисления фрактальной размерности для V -ритмики музыки. Для вычисления были взяты фрагменты из произведений Вивальди и концерта №5 Паганини. Сегментация проводилась с помощью той же программы, которая использовалась для исследования речи. Полученные результаты представлены на рис. 1.

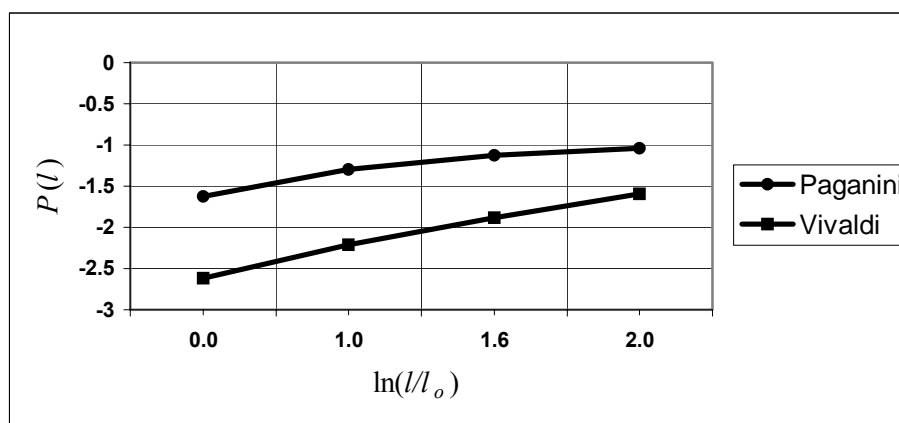


Рис. 1. Сравнение фрактальных размерностей V -ритмик музыки Вивальди и Паганини. Наибольшая нижняя граница размерности D_2 определяется как тангенс угла наклона

линейной регрессии следующих точек $\left(P(l) = \ln \frac{1}{N^2} \left(\sum_{i=1}^{M(l)} N_i^2 \right); \ln(l/l_0) \right)$, вычисленных при разных l

Как видно из рисунка, наклон линейной регрессии, или фрактальная размерность V-ритмики музыки Вивальди, равна $D_2 \approx 0.5$, что больше соответствует V-ритмике речи с упрощенным, фрактальным ритмом, свойственным речи заикающихся. Фрактальная размерность для фрагмента из музыки Паганини оказалась еще меньшей: $D_2 \approx 0.3$, т. е. V-ритмика музыки Паганини отстоит еще дальше от границы хаоса, чем музыка Вивальди, что соответствует более упрощенной ритмике данного фрагмента.

3.2. Колмогоровская энтропия или (в одномерном случае) показатель Ляпунова

Обобщим понятие информационной энтропии (энтропии Шеннона). По определению, энтропия Шеннона есть информация, которую мы получаем, узнав, что контрольная точка x_0 находится в конкретном из n подинтервалов, на которые разбит единичный отрезок. Для этой информации мы имеем следующую формулу:

$$I_0 = - \sum_{i=1}^n \frac{1}{n} \log_2 \frac{1}{n} = \log_2 n. \tag{5}$$

Обобщим эту формулу на рассматриваемый случай речевого «V»-ритма. Как оказалось, в случае «V»-ритма это обобщение допускает возможность экспериментальной оценки информационной энтропии. Рассмотрим отрезок ритма $\mathbf{T} = (T_1, \dots, T_N)$ (V-элементы ритма упорядочены членами натурального ряда). Будем следить за состоянием системы по мере разворачивания ритма. Пусть P_{i_1, \dots, i_n} есть совместная вероятность того, что элемент ритма T_1 (для определенности, его правая граница) находится в ячейке с номером i_1 , T_2 находится в ячейке с номером i_2 и так далее.

По аналогии с формулой (5), узнав, что ритмика системы находится в конкретном разбиении i_1^*, \dots, i_n^* с ячейками длиной l , мы, при условии знания априорных вероятностей P_{i_1, \dots, i_n} , получаем информацию:

$$I_n = - \sum_{i_1, \dots, i_n} P_{i_1, \dots, i_n} \ln P_{i_1, \dots, i_n}. \tag{6}$$

По аналогии с одномерным случаем, определим величину информации, диссипируемую системой генерации ритма при $n+1$ -ой итерации как $I_n - I_{n+1}$. Тогда К-энтропию Колмогорова для ритмики можно определить как среднюю величину информации, генерируемой системой за одну итерацию:

$$K = \lim_{l \rightarrow 0} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^{N-1} (I_n - I_{n+1}) = \lim_{l \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i_1, \dots, i_n} P_{i_1, \dots, i_n} \ln P_{i_1, \dots, i_n}. \tag{7}$$

Для удобства оценивания энтропии по экспериментальным данным вводят обобщенную конструкцию вида (так называемую энтропию Реньи):

$$K_f = -\lim_{l \rightarrow 0} \lim_{n \rightarrow \infty} \left(\frac{1}{n}\right) \left(\frac{1}{f-1}\right) \ln \sum_{i_1, \dots, i_n} P_{i_1, \dots, i_n}^f. \tag{8}$$

Нетрудно видеть с помощью правила Лопиталья, условия нормировки для вероятностей, а также с помощью факта монотонного убывания энтропия Реньи, как функции f , что справедливы следующие соотношения:

$$K_1 = K \text{ и } K_{f'} \leq K_f \text{ при } f \leq f'. \tag{9}$$

Особо выделяется K_2 как нижняя граница энтропии Колмогорова, причем эта граница зачастую достаточно проста для вычислений (во всяком случае, в случае одномерной ритмики речи), так как легко проверить выполнение следующего равенства:

$$K_2 = -\lim_{l \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \ln \sum_{i_1, \dots, i_n} P_{i_1, \dots, i_n}^2 = -\lim_{l \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \ln C_n(l). \tag{10}$$

Причем, в качестве корреляционного интеграла $C_n(l)$ может служить в случае речевой ритмики следующая величина, являющаяся обобщением определения корреляционного интеграла (N — общее число элементов ритмики) [5]:

$$C_n(l) = \lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{i=1, j=1}^N \theta(l - |\mathbf{T}_i^{(n)} - \mathbf{T}_j^{(n)}|). \tag{11}$$

Для прояснения термина «корреляционный» в определении корреляционного интеграла рассмотрим колебательную систему с аттрактором типа предельного цикла размерностью 1. Любая пара точек на этом предельном цикле, будет, очевидно, демонстрировать сильную корреляцию. С другой стороны, если пара точек относится к процессу с хаотической динамикой, то тогда, по определению, интервал между этими точками должен увеличиваться экспоненциально со временем из-за положительности показателя Ляпунова. Естественно, корреляция такой пары точек должна быть значительно меньше, чем в случае периодического движения.

В этом случае отклонение $C_n(l)$ от нуля будет служить уже мерой влияния вектора элементов ритмики $\mathbf{T}_i^{(n)} = (T_i, T_{i+1}, \dots, T_{i+n})$ на остальные векторы $\mathbf{T}_j^{(n)} = (T_j, T_{j+1}, \dots, T_{j+n})$ ($i \neq j$), а не одиночного элемента ритмики на остальные элементы, как предусматривалось обычным определением.

Так как физический смысл обобщенного корреляционного интеграла $C_n(l)$ допустимо трактовать (в соответствии с формулой для корреляционной размерности $R = \lim_{l \rightarrow 0} \left[\frac{\ln[C(l)]}{\ln l} \right]$, получаемой по аналогии с формулой для хаусдорфовой размерности) как обобщение на n -мерный случай соотношения $C(l) \sim (l)^R$, где R — корреляционная

размерность области фазового пространства соответствующей ритмики речи, то есть, как $C_n(l) \sim (l)^{nR}$, то допустимо и выполнение следующего равенства:

$$C_{n+1}(l) = C_n^{n+1}(l). \tag{12}$$

Используя прологарифмированное равенство (12), равенство (10) можно переписать в следующем виде:

$$K_2 = -\lim_{l \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \ln C_n(l) = \lim_{l \rightarrow 0} \lim_{n \rightarrow \infty} \ln \left[\frac{C_n(l)}{C_{n+1}(l)} \right]. \tag{13}$$

В случае речевой ритмики выражение (13) с помощью формулы (11) допускает построение довольно простого алгоритма компьютерного вычисления нижней границы энтропии Колмогорова K_2 из экспериментальных данных. Фактически формула (13) совместно с формулой (11) в случае конечной экспериментальной выборки представляет собой сжатую формулировку алгоритма. На рис. 2а и 2б приведены типичные результаты для оценки энтропии Колмогорова, получаемые нами для V-ритмики нормальной речи (рис. 2а) и участков речи, содержащих только нарушения плавности, возникающие в результате заикания (рис. 2б).

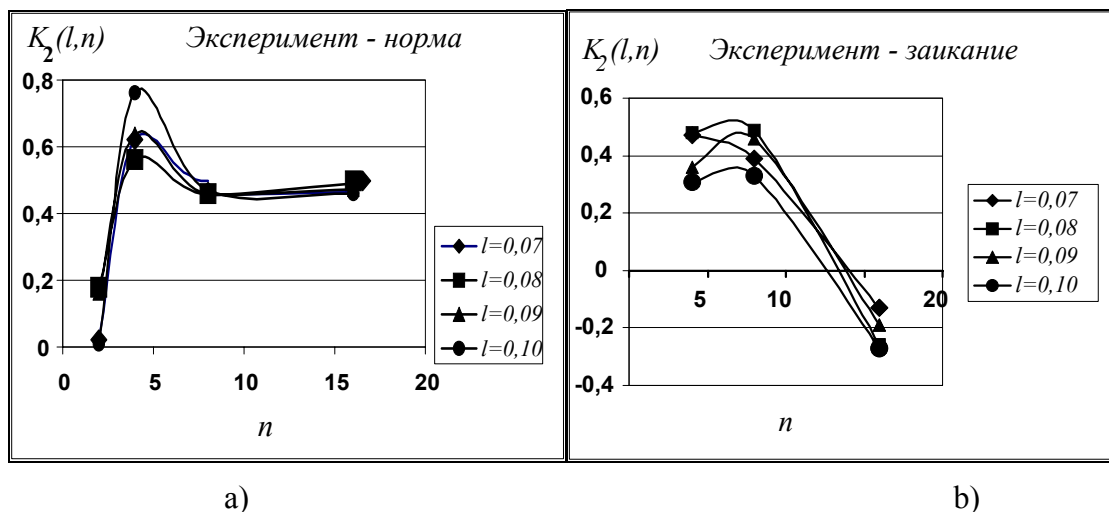


Рис. 2. Оценки снизу для энтропии Колмогорова для различных n и l

Из рисунка видно, что при $n \rightarrow \infty$ и $l \rightarrow 0$ энтропия Колмогорова для нормальной речи положительна и конечна. Это соответствует детерминированной хаотичности ритмики плавной речи. Для участков речи с нарушенным ритмом нижняя граница энтропии Колмогорова отрицательна, что, в принципе, допускает нулевое значение энтропии Колмогорова, а тем самым, допускается наличие периодических форм движения с нулевой информационной энтропией. Поскольку энтропия Колмогорова представляет собой информацию, генерируемую системой, то приведенные выше рассуждения позволяют характеризовать хаос, как состояние, в котором система «генерирует», или «диссипирует» информацию.

Таким образом, анализ структуры сегментированной речи показывает, что эта структура имеет хаотическую природу для нормальной речи и квазипериодическую фрактальную структуру для случая нарушенного ритма. Это заключение заставляет нас искать такую математическую модель процесса генерации ритма, которая, во-первых, имела бы как хаотический, так и фрактальный квазипериодический динамический режим.

4. МОДЕЛЬ РЕЧЕВОГО V-РИТМА

Во-первых, для того, чтобы удовлетворить первому условию, вспомним, что хаусдорфова размерность есть $D_2 = 0,75$ для ритма при заикании и $D_2 = 0,85$ для ритма нормальной речи. Таким образом, выполняется довольно хорошее согласие наших результатов с теоретическими оценками хаусдорфовых размерностей для логистического отображения $y_{n+1} = ry_n(1 - y_n)$; ($y_n \in [0,1]$), найденных в [6, стр. 121] для зоны бифуркаций и для зоны хаоса соответственно. Можно также показать, что определенные по таким размерностям индексы критичности задачи согласуются с постоянными Фейгенбаума для логистического уравнения. Кроме того, именно такие области наилучшей корреляции теоретического и экспериментального ритма мы наблюдали в зоне хаоса для нормальной речи, в то время как профиль такой корреляции при нарушенном ритме ярко выражен всплеском в бифуркационной зоне, рис. 3).

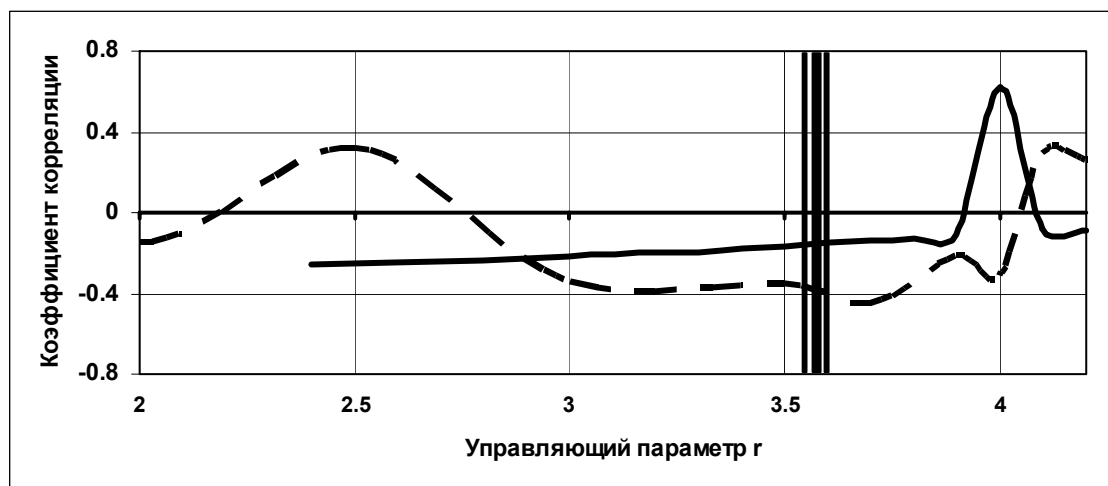


Рис. 3. Сравнение профилей коэффициента корреляции модели и экспериментальных данных на материале первой синтагмы.

Сплошная кривая — речевая фонация испытуемого с нормальной речью (К.Л.);
 пунктирная кривая — речевая фонация испытуемого, страдающего заиканием (В.В.);
 тройная вертикальная — граница между зоной хаоса (справа) и зоной бифуркаций (слева)

ЗАКЛЮЧЕНИЕ

В работе экспериментально установлено, что длительности сегментов речи генерируются для нормальной речи в фрактальном иррегулярном, хаотическом режиме, а при нарушении ритма вследствие заикания — в фрактальном квазиосцилляционном режиме. Экспериментальное исследование фрактальной размерности музыки показало, что ритм музыки соответствует формам ритма, характерным для начального этапа становления фонетики у детей. Можно предположить, что именно в этом кроются причины того, что восприятие как музыки, так и примитивной фонетики младенцев не зависит от языка, на котором говорит слушатель. Строение примитивных ритмов позволяет воспринимать звуки лишь эмоционально, но не на семантическом уровне. И только тогда, когда ритм приобретает иррегулярные черты, т. е. черты детерминированного хаоса, возможно семантическое, специфически языковое восприятие акустических ритмических последовательностей.

ЛИТЕРАТУРА

- [1] Feigenbaum M. J., Kadanoff L. P., Shenker S. J. Quasi periodicity in dissipate systems: renormalisation group approach. *Physica*, 1983, vol. 5D, 239–248.
- [2] Скляр О. П. Элементы теории ритма речи на основе физической феноменологии его нарушений. Автореф. дисс... доктора физ.-мат. наук. Санкт-Петербургский Университет, СПб, 1999.
- [3] Skljarov O. P. Neurodynamical Route to Chaos and Normal Speech vs. Stuttering. *Proceedings of Int. Conf. "Control of Oscillations and Chaos"*, St. Petersburg, 2000, 449–452.
- [4] Заславский Г. М., Сагдеев Р. З. Введение в нелинейную физику: от маятника до турбулентности и хаоса. М.: «Наука», 1988.
- [5] Шустер Г. Детерминированный хаос. М., «Мир», 1988.
- [6] Скляр О. П. V/U-ритм речи при чтении как индикатор состояния функции речевого дыхания у заикающихся. *Электронный журнал «Техническая акустика»*, <http://webcenter.ru/~eeaa/ejta>, 2004, 16.
- [7] Grassberger P., Procaccia I. On the Characterization of Strange Attractors, *Phys. Rev. Lett.*, 1983, vol. 50, 346–354.
- [8] Скляр О. П. Онтогенез речи и сценарий развития ее V-ритмов. *Электронный журнал «Техническая акустика»*, <http://webcenter.ru/~eeaa/ejta>, 2004, 7.