Детерминированный хаос

Изв. вузов «ПНД», т. 12, № 6, 2004 УДК 621. 391

О ХАОТИЧЕСКОЙ ПРИРОДЕ ЗВУКОВ РЕЧИ*

Ю.В. Андреев, М.В. Коротеев

Рассматриваются отдельные фонемы с точки зрения нелинейной динамики. Проводится анализ фазовых портретов сигналов в пространстве вложения, оценка размерности и старшего показателя Ляпунова. Показано, что речевые сигналы имеют невысокую размерность и положительный старший показатель Ляпунова.

Введение

В последнее десятилетие наблюдается определенный подъем интереса к изучению процесса речеобразования. Это связано как с практическими задачами сжатия информации, обусловленными возрастающей ролью коммуникаций в обществе, так и с развитием новых методов исследования сложных систем.

В предшествующий период основные успехи в исследовании речи, особенно в практическом плане, были достигнуты с помощью метода линейного предсказания [1, 2]. Он основан на том, что изменения речевого аппарата происходят на временах значительно больших, чем характерный период колебаний речевого сигнала. Поэтому на малых временных масштабах речевой аппарат можно считать стационарной системой и заменять его линейным фильтром, возбуждаемым специальным сигналом.

Хотя этот подход и основан на методе линейного предсказания, его название не должно вводить в заблуждение: неявно он предполагает нелинейную природу процесса речеобразования. С одной стороны, предсказание следующего отсчета дискретного речевого сигнала осуществляется при помощи линейной комбинации нескольких предыдущих отсчетов, однако с другой стороны, сигнал возбуждения

*Статья написана по материалам доклада на VII Международной школе «Хаотические автоколебания и образование структур», 1-6 октября 2004, Саратов, Россия

является в этой модели сложным сигналом с бесконечно широким спектром – периодической последовательностью импульсов для вокализованных звуков или шумовым сигналом для невокализованных. В то же время вопрос о природе источника, способного генерировать как широкополосные периодические сигналы, так и шумоподобные сигналы практически не обсуждается в литературе.

Помимо линейного предсказания, развивался также подход, связанный с физическим моделированием процесса речеобразования. Например, для вокализованных звуков была предложена нелинейная модель [3]. В ней голосовые связки представлены колеблющимися массами, а речевой тракт – двусторонней линией передачи (длинной линией). Нелинейным элементом является жесткость связи двух колеблющихся масс. Она представлена кусочно-кубической разрывной функцией, причем разрыв соответствует смыканию связок. При расчетах эту функцию аппроксимируют кусочно-линейной функцией. Модель описывается системой дифференциальных уравнений. Решения этой системы для разных наборов значений параметров довольно реалистично аппроксимируют некоторые гласные звуки [3].

С развитием теории динамического хаоса и появлением методов анализа хаотических систем, в том числе методов решения обратных задач нелинейной динамики, например, реконструкции аттрактора по наблюдениям одной переменной [4-6], методов реконструкции уравнений динамических систем [7] и др., появляются работы, в которых делаются попытки применить идеологию и методы нелинейной динамики к исследованию речевых сигналов. Среди них следует отметить работы [8, 9], в которых исследовались переходные явления в некоторых специфических звуках, таких как крики обезьян и плач младенцев. В них были обнаружены бифуркации удвоения периода, торы, переход к хаосу.

Наблюдавшиеся различными исследователями явления указывают на определенное сходство динамической системы речеобразования с поведением модельных хаотических систем. Поэтому имеет смысл попытка по наблюдаемому речевому сигналу, который в реальной системе может быть представлен, например, напряжением на выходе микрофона, восстановить неизвестную, вообще говоря, многомерную динамическую систему, его порождающую, или, по меньшей мере, оценить ее характеристики. В работах [10, 11] были выполнены такие виды исследований речевого сигнала, как восстановление фазового портрета, построение сечения Пуанкаре, оценка размерности фазового пространства соответствующей динамической системы, оценка старшего показателя Ляпунова. Исследования выполнялись с короткими, относительно стационарными фрагментами японской гласной фонемы «а» (часть исследований была проведена для гласной фонемы «и»).

В данной работе динамические характеристики речевых сигналов исследуются на более широком материале, включающем помимо основных гласных звуков некоторые фрикативные согласные (шипящие и свистящие). Заметим, что вообще нестационарность процессов речеобразования сильно усложняет применение к ним методов нелинейной динамики. В связи с этим, первым этапом должно стать изучение отдельных гласных и согласных звуков, а затем уже возможен переход к связной речи. Несмотря на трудности в изучении речевой динамики, вызываемые нестационарностью, именно она тесно связана с передачей информации в речевом сигнале. Стационарный квазипериодический или шумовой сигнал несет ничтожную информацию по сравнению с нестационарным речевым сигналом. Информативность речи

обусловлена ее нестационарностью, которая является следствием постоянных изменений в структуре речевого аппарата при формировании речи, в то время как при произнесении отдельных фонем речевой аппарат малоподвижен.

Для подтверждения хаотической природы речевых сигналов производилось восстановление динамической системы, соответствующей речевому сигналу, оценка корреляционной размерности, оценка старшего показателя Ляпунова. Обсуждаются трудности, возникающие при применении методов нелинейной динамики к речевым сигналам, и проблемы интерпретации полученных результатов.

1. Реконструкция фазового пространства речевых сигналов

При изучении временных рядов, в частности, дискретных речевых сигналов, применяют методы, развитые в последние десятилетия, такие как оценка показателей Ляпунова, оценка корреляционной размерности, анализ главных компонент. Эти методы используются и в настоящей работе.

Но первой задачей, которую нужно решить, прежде чем станет возможным применение методов нелинейной динамики, является реконструкция множества, соответствующего речевому сигналу, из временного ряда.

Исходным материалом для анализа является аналоговый речевой сигнал. Чтобы не потерять ключевой информации нужно выбрать разумную величину частоты дискретизации. Часто считают, что полоса частот речевого сигнала невелика, и что для большинства применений ее можно ограничить величиной порядка 4 кГц. Например, в телефонии сигнал ограничивают полосой 300–3500 Гц, для многих приложений, связанных со сжатием речи, сигнал оцифровывают с частотой 8 кГц, что согласно теореме Котельникова ограничивает его полосой 4 кГц. Как показывают результаты данного исследования, для некоторых звуков, у которых имеется шумовой участок в диапазоне 3–10 кГц, этого недостаточно, хотя это может быть неважно для задач сжатия или распознавания. В данной работе гласные оцифровывались с частотой 11.025 кГц, а некоторые согласные звуки с частотой 44.1 кГц.

Были исследованы гласные звуки «а», «о», «и», «э», «у» и согласные фрикативные «ш», «с», «з». Брались звуки, как произносимые отдельно, так и выделенные из связной речи, произнесенные как женскими, так и мужскими голосами.

Длительность записанных звуков составляла несколько секунд, однако для анализа использовались фрагменты длительностью от 0.5 до 1 секунды. При этом из каждого ряда, представляющего тот или иной звук, выделялись участки по возможности большей стационарности. Дискретизированный речевой сигнал (временной ряд) рассматривался как сигнал, произведенный неизвестной динамической системой. Необходимо заметить, что все методы исследования нелинейных динамических систем разработаны для стационарных, эргодических систем. Очевидно, систему речеобразования нельзя считать стационарной, точно так же нельзя говорить и об «аттракторе» речевой системы. Именно поэтому здесь будем рассматривать только более или менее стационарные участки речевого сигнала.

Помимо нестационарности существенной трудностью при применении методов нелинейной динамики к речи является временная многомасштабность – наличие в речевом сигнале колебаний с частотами, отличающимися на порядки. Это проявляется в том, что значимые частоты в спектре речевого сигнала могут занимать

диапазон от сотен герц до нескольких килогерц, что сильно отличает их от сигналов модельных хаотических систем, таких, например, как система Рёсслера.

Наиболее часто для реконструкции фазового пространства и аттрактора динамической системы применяется метод, основанный на теореме Такенса. Суть метода состоит в геометрической интерпретации сигнала динамической системы в евклидовом пространстве заданной размерности d. Временной ряд (x_k), $k = 1, 2, \ldots n$, прореживается с шагом τ. Чтобы получить точки пространства вложения R^d, из прореженного временного ряда берут последовательно по d отсчетов, так что $\bar{x}_i = \left(x_i, x_{i+\tau}, K, x_{i+(d-1)\tau}\right)$ [4, 6, 12, 13]. Точкам арифметического пространства взаимно-однозначно ставятся в соответствие векторы соответствующего пространства R^d. Величина τ носит название лага или времени задержки [13, 14]. Выбору лага, который во многом определяет качество вложения, посвящено большое количество литературы [13, 14], однако смысл рекомендаций часто сводится к выбору 2–4 точек на квазипериод колебаний. Однако, как отмечалось выше, в речевых сигналах присутствуют колебания, периоды которых отличаются на порядки. Если ориентироваться на период быстрых колебаний, то в построенном d-мерном множестве будет плохо представлена медленная динамика. Чтобы ее учесть, придется делать размерность пространства чрезмерно большой. Если ориентироваться на медленные колебания, есть опасность совершенно упустить из виду быстрые движения. По всей видимости, для речевых сигналов вместо равномерного прореживания «по Такенсу» нужно применять неравномерное, когда векторы для пространства R^d набираются так, что в векторе присутствуют как близкие, так и далекие элементы временного ряда. Некоторые рекомендации по такой технике можно найти в [15].

В данной работе было использовано равномерное прореживание при вложении траектории в пространство R^d. Применение очень длинных отрезков траекторий (до сотен тыс. точек) позволяет строить «плотные» множества в R^d, что, по мнению авторов, уменьшает последствия, связанные с многомасштабностью динамики сигналов. Построенные точки определяют фазовую траекторию неизвестной динамической системы и являются фазовыми портретами соответствующих звуков речи.

Для характеризации звуковых сигналов мы используем также сонограммы. В сонограммах вдоль оси абсцисс откладывается время, вдоль оси ординат – частота, а яркость цвета пропорциональна величине энергии, приходящейся на данную частоту в данный момент времени. Таким образом, сонограмма показывает эволюцию спектра сигнала во времени.

На рис. 1 и 2 приводятся реализации и сонограммы для гласного звука «а» и для согласного «ш». Важно отметить, что гласные звуки демонстрируют выраженную квазипериодическую структуру, в то время как рассмотренные согласные звуки носят шумоподобный характер, что видно как в реализациях, так и на сонограммах этих звуков. У гласных, как это показано на примере звука «а», энергия сконцентрирована в области низких частот и спектр имеет выраженные пики на основной частоте и формантных частотах (гармониках), что можно увидеть на сонограмме для этого звука в виде светлых полос, соответствующих большей энергии на данных частотах, в то время как для звука «ш» энергия гораздо более равномерно распределена по спектру, и сам спектр более соответствует спектру шума, то есть не содержит выраженных пиков. Похожая картина наблюдается и для других рассмотренных согласных звуков («с», «ф»).

При построении фазовых портретов, соответствующих звуковым сигналам, проводилась оценка размерности подпространства, занимаемого множеством, соответствующим звуку. Для этого использовалась процедура Брумхеда – Кинга (Карунена – Лоэва) [12]. Суть ее состоит в том, что в пространстве R^d для этого множества строится новый ортонормированный базис, оптимальный в смысле среднеквадратичной ошибки приближения этого множества. Он строится из собственных векторов ковариационной матрицы векторов множества, причем сумма квадратов проекций множества на оси нового базиса, то есть, фактически, энергия, приходящаяся на то или иное направление, определяется собственными значениями ковариационной матрицы. Это означает, что если множество занимает в пространстве подпространство меньшей размерности, то часть собственных значений будет нулевой или близкой к нулю, если исходные данные слегка зашумлены.

Уточняя описание процедуры Брумхеда – Кинга, мы можем сказать, что из векторов $\bar{x}_i$ строится матрица $X = \left(\bar{x}_1^T \bar{x}_2^T \dots \bar{x}_N^T\right)$, в которой исходные векторы представлены ее столбцами, затем находится ковариационная матрица $A = X \cdot X^T$ размера $d \times d$, и вычисляются ее собственные значения и собственные векторы. Затем множество проецируется на базис из собственных векторов [16].

Приведем фазовые портреты для двух звуков «а», произнесенных одним и тем же диктором в разных отрывках связного текста. Множества построены в трехмер-

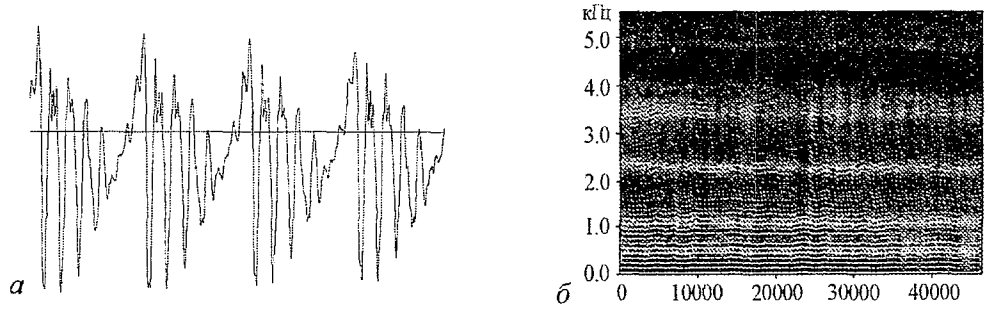

Рис. 1. Реализация и сонограмма для звука «а»

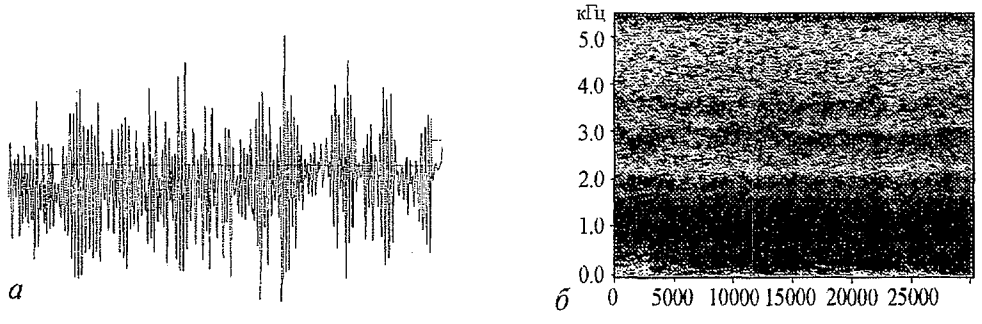

Рис. 2. Реализация и сонограмма для звука «ш»

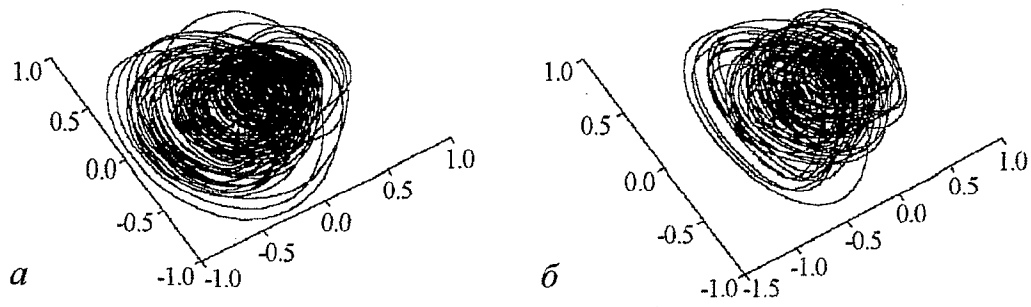

Рис. 3. Фазовые портреты звуков «а» из связных отрывков текста одного диктора

ном пространстве, однако для удобства сопоставления мы приводим их проекции на одну и ту же плоскость (рис. 3). Кроме того, приведем графики нормированных собственных чисел, определяющих размерность подпространства, в котором лежат множества (рис. 4). Они показывают, что основная часть энергии сигнала приходится на два первых собственных числа, в то время как уровень энергии, соответствующий другим собственным числам, значительно ниже. Видно также, что четвертое собственное число находится уже вблизи уровня 10^{-2}. Таким образом, с указанной точностью можно ограничиться трехмерным подпространством, определяемым тремя собственными числами, наибольшими по величине.

В целом, фазовые портреты звуков одного диктора демонстрируют некоторую схожесть, что отмечалось и для других гласных звуков. Кроме того, из приведенных фазовых портретов видно также, что они отличаются значительной гладкостью, что указывает на отсутствие значимых высокочастотных компонент, что свойственно гласным звукам.

Некоторое небольшое различие в динамике одного и того же звука можно объяснить редукцией гласного «а» в разных отрывках, что соответствует различию фонем звука «а» для данного диктора.

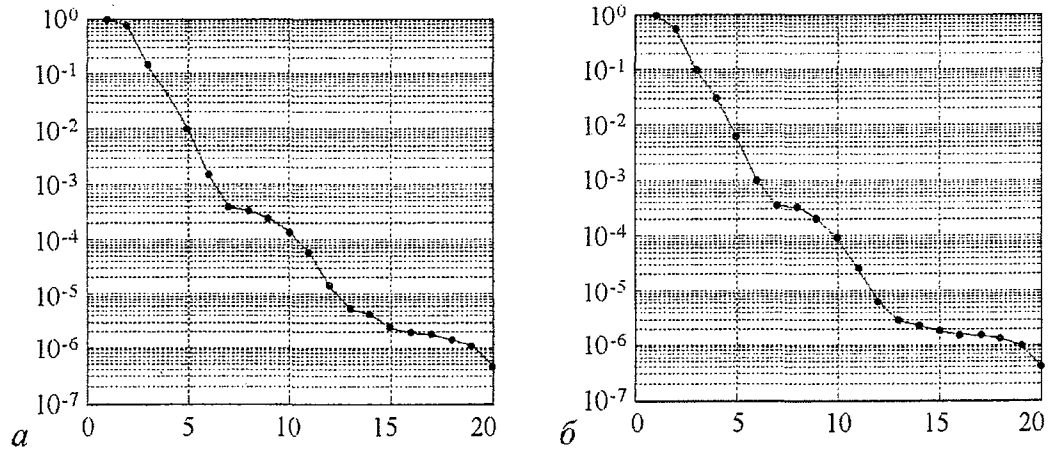

Рис. 4. Собственные числа в порядке убывания их величины для множеств на рис. 3

Аналогичное рассмотрение было проведено и для других гласных звуков. Можно сказать, что фазовые портреты гласных звуков имеют, как правило, схожие особенности у одного и того же диктора. Кроме того, структура фазовых портретов подтверждает «почти» периодический характер гласных звуков, осложняемый некоторой нестационарностью. Фазовые портреты гласных звуков отличаются гладкостью и повторяемостью. Оценки размерности подпространства фазового пространства, в котором располагаются множества, соответствующие гласным звукам, полученные с помощью спектра собственных значений ковариационной матрицы, указывают, что размерность этого множества невелика и находится на уровне 3–4. Для всех гласных звуков можно отметить также присутствие области повышенной мощности в спектре на уровне 9–13 кГц, причем устранение этой области слабо сказывается на качестве речи. В спектрах, вполне аналогично тому, как это было показано на примере звука «а», четко выделяется основная частота и формантные частоты (гармоники). При исследовании отдельно произносимых гласных звуков замечено, что в силу условий формирования они получаются более стационарными, а фазовые портреты отличаются гладкостью по сравнению со звуками, выделенными из связной речи.

2. Двойственная природа звука «з»

Исследование звука звонкой фрикативной согласной «з» выявило некоторые черты, которые роднят его как с гласными звуками, так и с шумовыми шипящими. Фазовый портрет звука «з» (женский голос) приведен на рис. 5, *а*. Так как это не гласный звук, можно было бы ожидать существенного отличия структуры его фазового портрета от портретов гласных звуков. Однако его фазовый портрет имеет общие черты с портретами гласных. В частности, в нем хорошо заметна тороидальная структура. В то же время, наблюдается высокочастотная изрезанность, которая свидетельствует о наличии высокочастотной шумовой компоненты в сигнале. Эта компонента отчетливо наблюдается на спектре мощности сигнала (рис. 6). По уровню спектральной плотности она примерно на 30 дБ ниже, чем главная низкочастотная компонента. Однако эта компонента сигнала занимает полосу порядка 2000 Гц, а узкополосная низкочастотная, регулярная компонента – только 10 Гц. Поэтому их интегральные мощности отличаются примерно в десять раз. Это позволяет наблюдать присутствие шумовой компоненты сигнала в фазовых портретах.

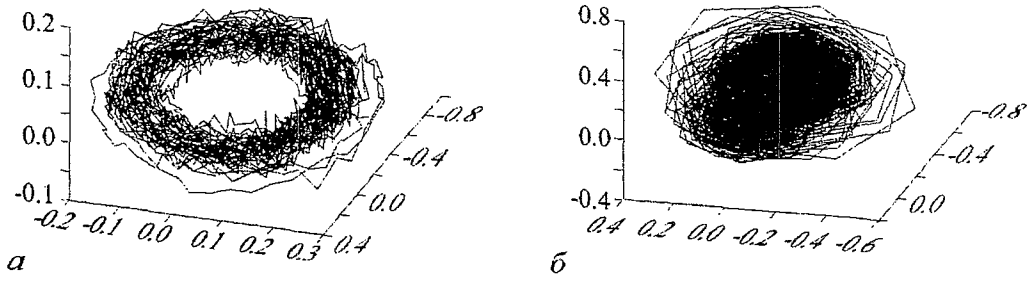

Рис. 5. Фазовый портрет для главных компонент звука «з»: *а* – женский голос, *б* – мужской голос

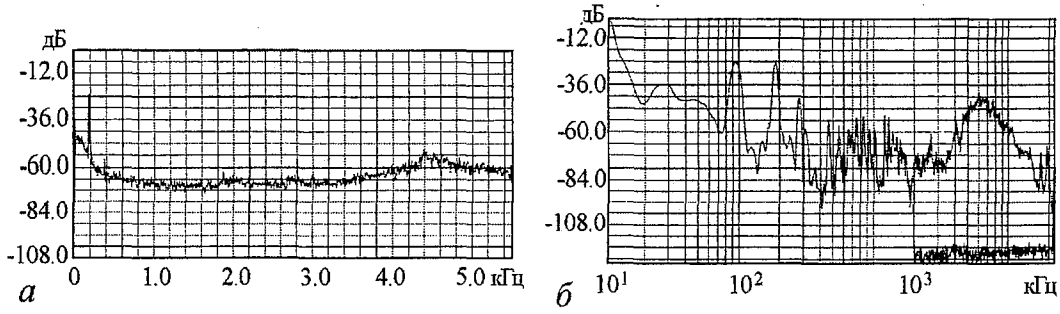

Рис. 6. Спектр мощности звука «з»: *а* – женский голос, *б* – мужской голос (нижняя кривая – спектр сигнала после ФНЧ 0–2000 Гц)

Анализ структуры фазового портрета и спектра звука «з», произносимого женским голосом, показывает, что частота 11 кГц для оцифровки, по-видимому, недостаточна. Поэтому сигнал звука «з» мужского голоса был оцифрован с частотой 44.1 кГц. Типичный фазовый портрет такого сигнала приведен на рис. 5, *б*. Его вид с характерными изломами показывает, что даже частоты дискретизации в 44.1 кГц недостаточно для получения гладких данных.

Спектральная структура звука «з» для мужского голоса отличается от соответствующей структуры спектра для женского голоса (рис. 6). В частности, этот спектр содержит большое число гармоник основной частоты. Как и спектр мощности для женского голоса он содержит шумовую компоненту в области частот от 3 до 12 кГц. Однако линейчатые спектральные компоненты простираются практически до «шумовой» области спектра.

Таким образом, звук «з» имеет двойственную структуру – он объединяет в себе свойства вокализованных сигналов (гласных звуков), что проявляется в наличии тороидальной структуры фазового портрета и присутствии основного тона и его гармоник в спектре, и шумоподобных сигналов, что добавляет шумовую высокочастотную компоненту как к портретам, так и к спектру. Чтобы проиллюстрировать сказанное, выделим из сигнала низкочастотную и высокочастотную составляющие.

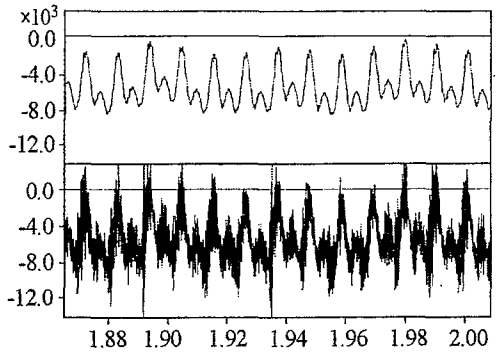

Рис. 7. Звук «з» – мужской голос, вверху – после применения ФНЧ 2 кГц, нижняя кривая – исходный сигнал

После применения ФНЧ с полосой 0–2000 Гц получаем сигнал на рис. 7 (сверху). Спектр этого сигнала приведен на рис. 6, *б* (нижняя кривая). Даже визуально сигнал очень напоминает сигналы гласных звуков. И действительно, на слух такой сигнал воспринимается как вокализованный звук, напоминающий «у» или «ы».

Как следует из приведенных фазовых портретов низкочастотной составляющей сигнала «з», этот сигнал имеет выраженную тороидальную структуру довольно низкой размерности. Как видно из проекции портрета на коорди-

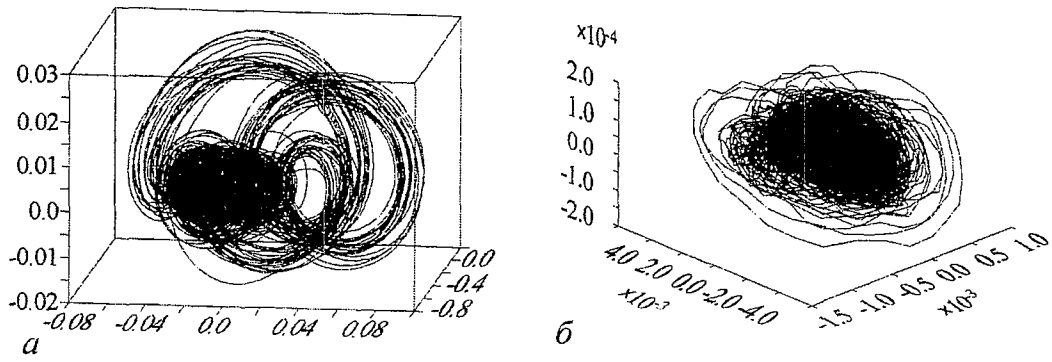

Рис. 8. Проекции фазового портрета в базисе собственных векторов отфильтрованного звука «з» на координаты 1–2–3 (*а*) и 4–5–6 (*б*)

наты 4–5–6 (рис. 8), она представляет собой шумовой «остаток» существенно меньшей амплитуды. Низкая размерность полученного множества подтверждается расчетом спектра собственных значений ковариационной матрицы, дающей оценку размерности множества $d = 3$.

Для того чтобы выяснить, чему соответствует высокочастотная область «шума» на спектре исходного звука, исходный сигнал был пропущен через полосовой фильтр 3–10 кГц. Результаты указанной фильтрации приводятся на рис. 9.

На слух отфильтрованный в полосе 3–10 кГц сигнал воспринимается как шипящий, близкий к «с». Как следует из анализа спектра собственных значений, размерность «шумовой» составляющей сигнала «з» также ограничена небольшой величиной $d = 4$–5, что подтверждается и проекциями фазового портрета (рис. 10). Все это свидетельствует в пользу ее динамического происхождения.

Таким образом, звук «з» действительно сочетает в себе черты гласного и согласного звуков, причем обе его составляющие, по-видимому, являются достаточно низкоразмерными. Остается вопрос, каким образом формируется звук такой двойственной природы? Авторы предполагают, что сигнал на рис. 6 может быть по-

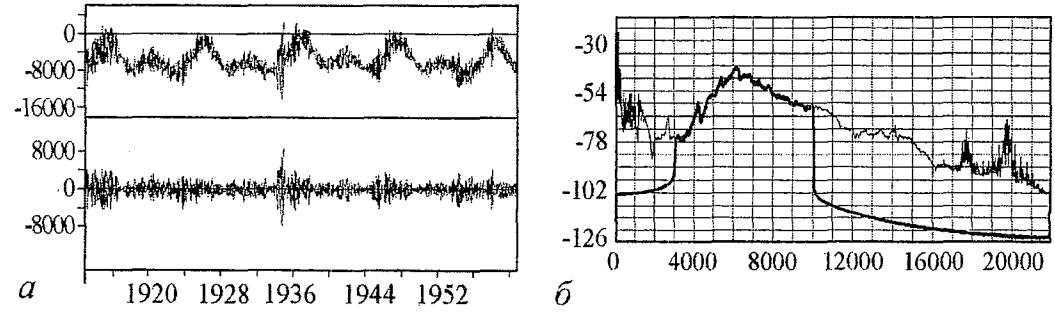

Рис. 9. Характеристики «шумовой» части звука «з»: *а* – реализация исходного сигнала (верхняя кривая) и отфильтрованного в полосе 3–10 кГц; *б* – спектр исходного сигнала и отфильтрованного

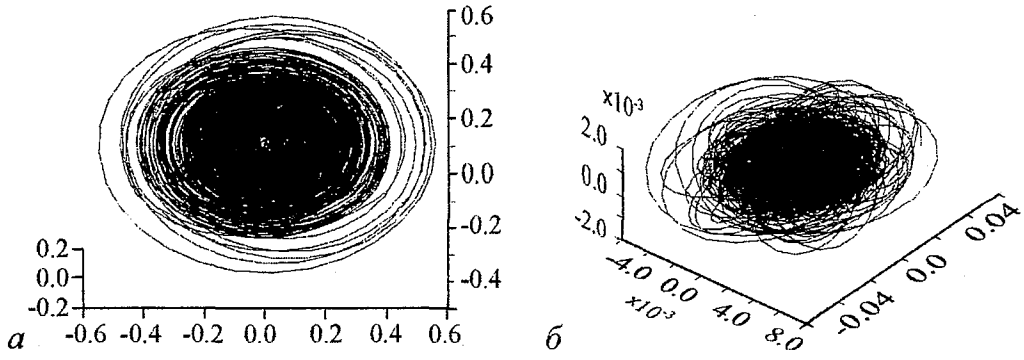

Рис. 10. Фазовый портрет отфильтрованного (3–10 кГц) сигнала в базисе собственных векторов. *а* – координаты 1–2–3, *б* – координаты 4–5–6

рожден путем возбуждения нелинейного высокочастотного резонатора сравнительно низкочастотным гармоническим сигналом, однако подтверждение этой гипотезы требует дальнейшего исследования.

3. Оценка старшего ляпуновского показателя

Как известно, D-мерная динамическая система описывается набором d показателей Ляпунова. Если динамический режим системы является хаотическим, то старший показатель Ляпунова (максимальный) положителен. Это означает, что две траектории, выходящие из малой окрестности, экспоненциально быстро разбегаются. В простых модельных системах, как правило, только старший показатель Ляпунова является положительным, остальные – отрицательны. В настоящей работе мы приводим оценки старшего ляпуновского показателя для различных звуков, как выделенных из отрывков текста, так и отдельно произносимых.

Вычисление старшего показателя Ляпунова несложно, если известны уравнения движения динамической системы. В этом случае для точек на траектории системы вычисляют локальные коэффициенты растяжения, которые затем усредняют, получая, таким образом, старший показатель Ляпунова. Для получения локального коэффициента растяжения в момент времени t_k в окрестности текущей точки траектории $\mathbf{x}_k$ берут близкую точку $\mathbf{x}'_k$ и на отрезке времени Δt помимо основной траектории рассчитывают траекторию, исходящую из точки $\mathbf{x}'_k$. Локальный показатель Ляпунова рассчитывается по формуле

$$\lambda_k = \frac{1}{\Delta t} \ln \left(\frac{\|\mathbf{x}_{k+1} - \mathbf{x'}_{k+1}\|}{\|\mathbf{x}_k - \mathbf{x'}_k\|} \right),$$

где $\mathbf{x'}_{k+1}$ и $\mathbf{x}_{k+1}$ – точки соответствующих траекторий в момент времени $(t_k + \Delta t)$. Затем вектор $(\mathbf{x'}_{k+1} - \mathbf{x}_{k+1})$ нормируют, чтобы его длина оставалось малой, и повторяют процедуру.

Если уравнения движения не известны, а известна только одна наблюдаемая траектория системы, то расчет траектории, близкой к наблюдаемой, невозможен. В таком случае применяют локально-линейные предикторы для расчета траекторий, проходящих вблизи наблюдаемой траектории [17]. Если наблюдаемая траектория достаточно длинна, то она много раз пройдет по одной и той же области фазового пространства, и в окрестности любой ее точки найдется много других близких точек. Сначала находят векторы, попадающие в окрестность текущей точки. Затем строят линейное преобразование окрестности текущей точки в окрестность следующей точки на траектории (совместно для всех точек окрестности). На основе матрицы данного линейного преобразования вычисляют коэффициент растяжения вектора отклонения на временном интервале, соответствующем переходу от текущей точки траектории к следующей. Логарифм этого коэффициента является локальным показателем Ляпунова в текущей точке. В окрестности следующей точки процедуру повторяют. Показатель Ляпунова динамической системы получают усреднением локальных показателей на достаточно большом участке траектории [18, 19]. Очевидно, точность полученной оценки определяется линейностью преобразования окрестности в следующую окрестность, а значит, наличием достаточного количества близких точек в окрестности каждой точки траектории, что определяется длиной траектории.

Приведем оценки старшего ляпуновского показателя для отдельного гласного звука. Далее будут приведены результаты расчета для всех рассмотренных в работе звуков, однако более детальное изложение мы дадим для гласного звука «а», как и ранее.

На рис. 11 приводится отрывок реализации и восстановленный фазовый портрет для отдельно произнесенного звука «а». На рисунке фрагмента сигнала видны два временных масштаба – «большой период» $T_1 = 50$ отсчетов, соответствующий частоте основного тона 200 Гц (частота дискретизации 11025 Гц), и «период» малых колебаний $T_2 = 13$ (850 Гц). Фазовый портрет сигнала в пространстве вложения приводится на соседнем рисунке. Портрет зарисован двумя изначально очень близкими траекториями (начальное отклонение = 0.024). В совокупности эти две траектории, быстро разбежавшись, представляют сложный торообразный портрет этого сигнала.

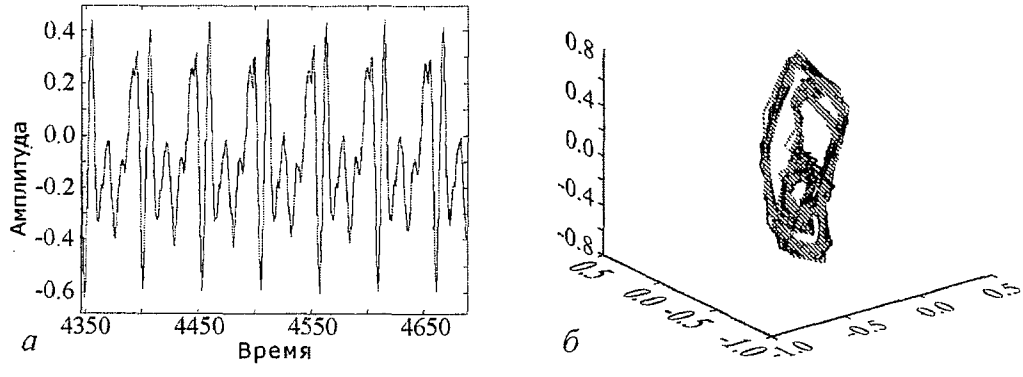

Рис. 11. Отрывок реализации гласного звука «а» и фазовый портрет этого звука (женский голос). На фазовом портрете показана исходная траектория и траектория локально-линейного предиктора

Таблица

Старший ляпуновский показатель для различных звуков при реализации мужскими и женскими голосами

звук	λ	Характеристики
И (ж)	5050 (8,4)	Fs = 11025 Гц 10000 отсчетов
А (м)	2830 (4,7)	
У (ж)	4750 (7,9)	20000 отсчетов
У (м)	732 (1,2)	
Э (ж)	3210 (5,3)	10000 отсчетов
Э (м)	3800 (6,3)	10000 отсчетов
О (ж)	2750 (4,5)	10000 отсчетов
О (м)	2160 (3,6)	10000 отсчетов
С (ж)	6360 (10,6)	10000 отсчетов
Ш (ж)	6140 (10,2)	20000 отсчетов
Ш (м)	2830 (4,7)	20000 отсчетов

Оценка показателя Ляпунова с помощью методики, описанной выше, для этого сигнала дает $\lambda \sim 600\,c^{-1}$. Для того чтобы можно было сравнивать показатели Ляпунова реальных сигналов с модельными системами такими, например, как система Лоренца, нужно привести их к одному временному масштабу. Так, в системе Лоренца средний период малого колебания равен ~ 0.7, а в нашем случае $\sim 13/11025$, то есть масштабы различаются приблизительно в 600 раз. В пересчете к временному масштабу системы Лоренца для звука «а» получаем $\lambda \sim 1$, то есть показатели Ляпунова обеих систем сопоставимы. Для гласных звуков «у» и «и» получены близкие значения оценок показателя Ляпунова в относительных величинах для сравнения с системой Лоренца. Существенно, однако, отметить другое. Положительность старшего ляпуновского показателя указывает на хаотический характер, присущий динамике речевого аппарата. В таблице приводятся образцы расчетов старшего ляпуновского показателя в различных гласных и согласных звуках при различном качестве стационарности рассматриваемого временного ряда. Мы даем абсолютные значения старшего показателя, а также для удобства сравнения в скобках приводим те же величины, пересчитанные в соответствии с временным масштабом системы Лоренца. Следует отметить значительное увеличение показателя Ляпунова для согласных звуков. Кроме того, как уже неоднократно упоминалось ранее, в силу нестационарности процесса речеобразования мы получаем различные значения для старшего показателя на стационарных участках и на больших участках произнесения звука, на которых уже сказываются нестационарные эффекты.

4. Вычисление корреляционной размерности речевых сигналов

Одним из основных методов оценки размерности множества, получающегося в пространстве вложения для речевого сигнала, является метод оценки корреляционной размерности. Здесь ситуация вполне аналогична той, которая имеет место

для оценки размерности аттрактора хаотической динамической системы. Для нахождения корреляционной размерности сначала вычисляют корреляционный интеграл [6, 14]

$$C(\varepsilon, N) = \frac{1}{N^2} \sum_{i \neq j} \theta(\varepsilon - |\bar{x}_i - \bar{x}_j|),$$

где $\bar{x}_i, \bar{x}_j$ – точки d-мерного пространства вложения, ε – расстояние между этими точками, а θ – функция Хевисайда. Тогда корреляционная размерность определяется из соотношения

$$\nu = \lim_{\varepsilon \to 0} \lim_{N \to \infty} \frac{\lg C(\varepsilon, N)}{\lg \varepsilon},$$

где N – число точек в фазовом пространстве вложения [20]. Корреляционная размерность также дает оценку размерности того множества, которое получается для речевых сигналов в фазовом пространстве [13, 20].

Приведем результаты, относящиеся к вычислению корреляционной размерности для различных гласных и согласных звуков, произносимых различными дикторами, а также размерности для небольших связных отрывков речи. На рис. 12 и 13 приводятся графики корреляционной размерности при различных размерностях пространства вложения для отдельно произносимых звуков.

Оценка корреляционной размерности дает значимый результат лишь в том случае, когда рост размерности пространства вложения приводит к определенному насыщению, то есть когда кривые, соответствующие различному выбору размерности пространства, начинают существенно приближаться друг к другу. Такая ситуация

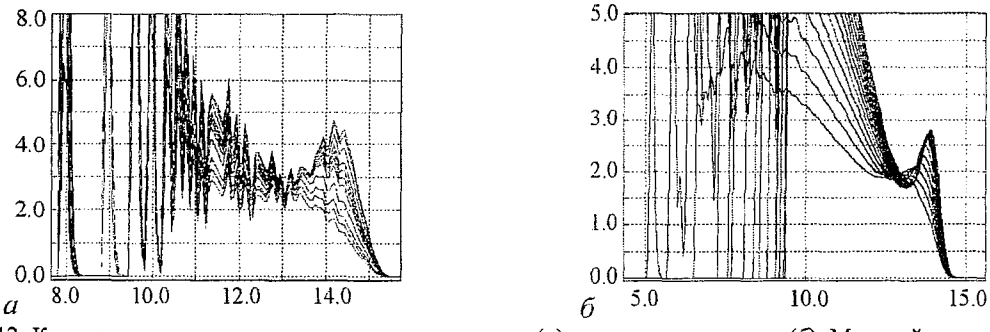

Рис. 12. Корреляционная размерность гласного звука «а» (*а*) и гласного звука «у» (*б*). Мужской голос. Размерности пространства вложения 4–18

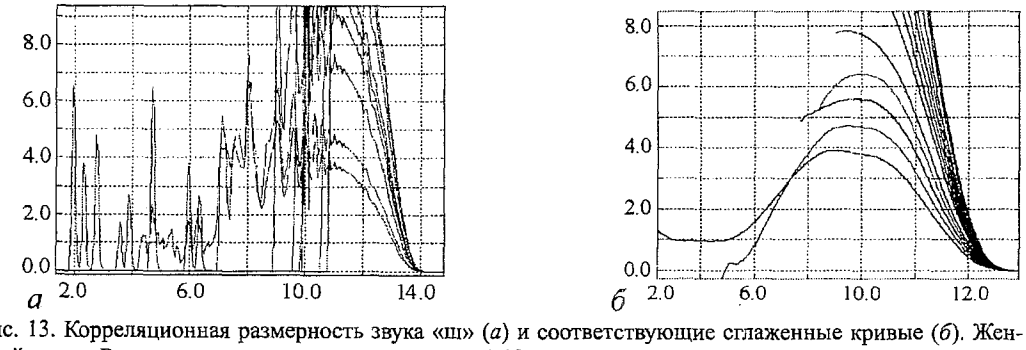

Рис. 13. Корреляционная размерность звука «ш» (*а*) и соответствующие сглаженные кривые (*б*). Женский голос. Размерности пространства вложения 4-18.

видна на рис. 12, *а* и *б*, где область насыщения дает величину корреляционной размерности 2.5–3 в обоих случаях. Шумоподобный звук «ш» дает иную картину (рис. 13). Для него не удается получить явного участка с насыщением, что указывает на увеличение размерности множества в этом случае. Впрочем, следует отметить, что указанное обстоятельство вполне согласуется с фазовыми портретами для звука «ш», которые указывали на то, что энергия в данном случае распределена по собственным значениям более равномерно, нежели в случае гласных звуков.

Выводы

В данной статье представлены результаты анализа применения методов нелинейной динамики для анализа речевых сигналов. В качестве результатов анализа мы приводим данные по различным гласным и согласным звукам, как специально записанным для данного исследования, так и выделенным из связной речи. В силу большого объема полученных результатов в качестве иллюстрации мы приводили, в основном, данные, полученные для гласного звука «а» и некоторых согласных. Нами были таким же образом проанализированы и другие гласные звуки: «и», «э», «о». Мы сознательно избегали пока анализа двугласных в силу их более сложной природы. Все гласные звуки, выделенные из речи, обладают в той или иной степени квазипериодической структурой, в отличие от согласных звуков, часть которых носит шумовой характер, а часть должна рассматриваться как сложные переходные процессы. Обобщая все полученные частные результаты, можно сказать следующее.

При рассмотрении речевых сигналов мы имеем дело со сложной нестационарной системой. Главной целью настоящей работой служило выявление нелинейного динамического характера системы, порождающей речевой сигнал, а также хаотическое поведение в этой системе при речеобразовании. Этот вопрос, как нам представляется, не может быть решен в одной публикации, потому что является чрезвычайно обширным. В работе мы ограничились лишь рассмотрением отдельных звуков, однако от звуков необходимо совершать переход к их сочетаниям, а затем и к связной речи.

Показано, что при вложении, согласно методике, основанной на теореме Такенса, в пространство большой размерности, гласные звуки можно с хорошей точностью выделить в нем как некоторые множества, лежащие в подпространстве небольшой размерности. Для всех гласных звуков размерность этого подпространства составляет примерно 3–4. Представленные в статье результаты по анализу отдельных гласных звуков укрепляют доказательную базу о превалировании нелинейных эффектов при формировании речевого сигнала.

Характерные особенности произнесения звуков, присущие данному диктору, обнаруживаются при рассмотрении восстановленных множеств в фазовом пространстве динамической системы в виде особой структуры, присущей построенным множествам. Таким образом, замечено, что особенности речеобразования конкретного человека можно, в принципе, трактовать геометрически, что способно открыть пути к созданию банков речевых звуков, где они трактуются с этой точки зрения и могут быть применены в различных областях.

Расчет старшего ляпуновского показателя, проведенный для всех рассмотренных звуков, и его положительная величина указывает на присутствие хаотического

поведения в динамической системе речевого тракта, что является существенным аргументом в пользу наличия хаотической нелинейной динамики в речеобразовании.

Проведенный расчет корреляционной размерности для гласных звуков дает оценку размерности пространства вложения различных звуков. Для всех гласных звуков оценка корреляционной размерности дает значение на уровне 3–4, в то время как для рассмотренных согласных звуков не удается получить четкого значения этой характеристики.

Библиографический список

1. *Маркел Дж., Грэй А.* Линейное предсказание речи. М.: Связь, 1980. 308 с.
2. *Макхол Дж.* Линейное предсказание. Обзор // ТИИЭР. 1975. Т. 53, № 2. С. 20.
3. *Ishizaka K., Flanagan J.L.* Synthesis of voiced sounds from a two-mass model of the vocal cords // The Bell System Technical Journal. July-August, 1972. Vol. 51, № 6. P. 1233.
4. *Takens F.* Detecting strange attractors in turbulence // Lecture notes in mathematics, № 898, Springer-Verlag, 1981. P. 366-381.
5. *Noakes L.* The Takens embedding theorem // Int. J. Bifurcation and Chaos. 1991. Vol. 1. P. 867.
6. *Farmer J.D., Sidorowich J.J.* Predicting chaotic time series // Phys. Rev. Lett. 1999. Vol. 59, № 8. P. 845.
7. *Безручко Б.П., Диканев Т.В., Смирнов Д.А.* Тестирование на однозначность и непрерывность при глобальной реконструкции модельных уравнений по временным рядам // Известия вузов. Прикладная нелинейная динамика. 2002. Т. 10, № 4. С. 69.
8. *Herzel H., Berry D., Titze I.R., Saleh M.* Analysis of vocal disorders with method from nonlinear dynamics // J. Speech Hear. Res. 1994. Vol. 37. P. 1008-1019.
9. *Titze I.R.* The physics of small-amplitude oscillation of the vocal folds // J. Acoust. Soc. Am. 1988. Vol. 83. P. 1536-1552.
10. *Tokuda I., Tokunaga R., Aihara K.* A simple geometrical structure underlying speech signals of the japanese vowel /a/ // Int. J. of Bifurcation and Chaos. 1996. № 1. P. 159.
11. *Tokuda I., Miyano T., Aihara K.* Surrogate analysis for detecting nonlinear dynamics in normal vowels // J. Acoust. Soc. Am. Dec., 2001. Vol. 110(6). P. 3207.
12. *Broomhead D.S., King G.P.* Extracting qualitative dynamics from experimental data // Phys. D. 1986. Vol. 20. P. 217.
13. *Kantz H., Schrider T.* Nonlinear time series analysis. Cambridge university press, 2000. 304 p.
14. *Lai Y.-C., Ye N.* Recent developments in chaotic time series analysis // Int. J. Bifurcation and Chaos. 2003. Vol. 13, № 6. P. 1383.
15. *Judd K., Mees A.* Embedding as a modeling problem // Physica D. 1998. Vol. 120. P. 273.
16. *Ланда П.С., Розенблюм М.Г.* Об одном методе оценки размерности вложения аттрактора по результатам эксперимента // ЖТФ. 1989. Т. 59, вып. 1. С. 13.

17. *Grassberger P., Procaccia I.* Characterization of strange attractors // Phys. Rev. Lett. 1983. Vol. 50. P. 346.
18. *Eckmann J.-P., Kamphorst S.O., Ruelle D., Ciliberto S.* Lyapunov exponents from time series // Phys. Rev. A. 1986. Vol. 34. P. 4971.
19. *Ланда П.С., Розенблюм М.Г.* Сравнение методов конструирования фазового пространства и определения размерности аттрактора по экспериментальным данным // ЖТФ. 1989. Т. 59, вып. 11. С. 1.
20. *Grassberger P., Procaccia I.* Measuring the strangeness of strange attractors // Physica D.1983. Vol. 9. P. 189.

Институт радиотехники и электроники РАН, Москва

Поступила в редакцию 20.12.2004

ON CHAOTIC NATURE OF SPEECH SIGNALS

Yu.V. Andreyev, M.V. Koroteyev

Phonetic signals are considered from the viewpoint of nonlinear dynamics. Phase portraits of the signals are analyzed in embedding space, dimension and the largest Lyapunov exponent are estimated. It is shown that dimension of speech signals is low and the largest Lyapunov exponent is positive.

Андреев Юрий Вениаминович – родился в Уфе (1960), окончил Московский Физико-технический институт (1983). Защитил диссертацию на соискание ученой степени кандидата физико-математических наук в Институте радиотехники и электроники РАН (1993) в области обработки информации методами нелинейной динамики. Опубликовал более 20 работ в этой и смежных областях. Старший научный сотрудник ИРЭ РАН.
E-mail:yuwa@cplire.ru

Коротеев Максим Валерьевич – родился в 1976 году. В 1999 году окончил факультет аэромеханики и летательной техники МФТИ. В настоящее время младший научный сотрудник Института радиотехники и электроники РАН. Кандидат физико-математических наук (2004). Область научных интересов – динамический хаос, анализ временных рядов, гидродинамика, асимптотические методы. Автор более 10 работ.