УДК 004.934.2

ПРИМЕНЕНИЕ МЕТОДОВ НЕЛИНЕЙНОЙ ДИНАМИКИ ДЛЯ РАСПОЗНАВАНИЯ ЭМОЦИИ РАДОСТИ В РЕЧИ

К.В. Сидоров, Н.Н. Филатова

Рассмотрена задача распознавания образцов речи, зарегистрированных в момент проявления испытуемыми эмоции радости, от образцов речи этих же дикторов в нейтральном состоянии. Для решения задачи использованы методы нелинейной динамики. Исследования проведены на записях, взятых из базы Emo-DB (Берлин), и фрагментах русскоязычной базы (Тверь). Сформирован модельный корпус эмоциональной речи, состоящий из базы данных двух уровней (фраз и фонем), послуживший основанием для оценки работоспособности разрабатываемых алгоритмов. Выделены устойчивые признаки нелинейной динамики – реконструкция аттрактора и рекуррентный график. Предложены новые количественные признаки для классификации образцов речи человека, испытывающего эмоцию радости, основанные на оценках максимальных векторов реконструкции аттрактора для четырех квадрантов.

Ключевые слова: эмоция, эмоциональное состояние, речь, речевой сигнал, нелинейная динамика, реконструкция аттрактора, рекуррентный график.

Введение

На современном этапе развития информационных технологий разработка методов и систем распознавания эмоционального состояния человека по речевому сигналу с помощью аппаратно-программных средств является актуальной задачей, позволяющей решить ряд проблем в области биомедицинских технологий. В последние годы наблюдается явное усиление интереса к анализу речевого сигнала как объективного показателя эмоционального состояния человека [1, 2]. Различные исследования в области акустики, психолингвистики и психофизиологии позволили собрать сведения о множестве акустических, просодических и лингвистических характеристик речи, которые можно использовать в качестве информативных признаков при распознавании эмоционального состояния, проявляющихся на уровне сегментов, фонем (звуков), слогов, целых слов и фраз. Чаще всего используются следующие признаки речевого сигнала [3]: спектрально-временные, амплитудно-частотные, вейвлет, кепстральные и характеристики (инварианты) нелинейной динамики. Судя по полученным результатам, перечисленные признаки зарекомендовали себя с положительной стороны. Однако, несмотря на большое количество проведенных в данном направлении исследований, ряд проблем все еще остается нерешенным, и многие идеи требуют дальнейшего развития. В частности, отсутствует универсальная теоретическая модель описания речевых образцов в условиях проявления разных видов эмоций, отражающая взаимосвязь вида эмоций и объективных характеристик речевого сигнала.

На текущий момент времени выделение новых информативных признаков, по возможности родственных человеческому восприятию, и поиск эффективных методик распознавания эмоций, являются важнейшей задачей. В работе рассматривается способ решения этой задачи методами нелинейной динамики, позволяющими получить количественную и качественную оценку признаков, проявляющихся в речевом сигнале человека, испытывающего эмоцию радости.

Модельный корпус эмоциональной речи

В настоящее время в Тверском государственном техническом университете активно ведутся разработки системы распознавания эмоционального состояния человека по образцам речевого сигнала. Для проведения исследований необходимо наличие модельного корпуса эмоциональной речи, т.е. базы дан-

ных, в которой хранятся образцы речи испытуемых, находящихся в различных эмоциональных состояниях. В связи с этим был сформирован модельный корпус эмоциональной речи, состоящий из двух частей (русской и немецкой). При создании русскоязычной части в качестве дикторов (испытуемых) выступили 5 человек, каждый из которых, на основе одного нейтрального образца, создал несколько клонов с различным уровнем проявления положительной эмоции радости [4], выбор которой обусловлен интересами дальнейшего применения разрабатываемой технологии. При формировании немецкоязычной части использовались записи эмоции радости и нейтрального состояния, взятые из берлинской базы данных эмоциональной речи Emo-DB (Berlin Database of Emotional Speech) [5], состоящей из 535 фраз 10 дикторов, имитирующих набор эмоциональных состояний: гнев, скука, отвращение, беспокойство/страх, печаль, радость/счастье и нейтральное состояние. В целом, модельный корпус состоит из двух уровней, связанных иерархически. Уровень 1 включает образцы фраз от разных дикторов. Используя алгоритм автоматической генерации речевых объектов [6] для каждой записи уровня 1, получены объекты уровня 2 – фонемы. Всего для проведения исследований сформированы 4 обучающие выборки (ОВ):

1. ОВ 1.1 – 18 русских записей контрольной фразы «А голос мой звучит примерно так»;
2. ОВ 1.2 – 180 гласных фонем, полученных из ОВ 1.1;
3. ОВ 2.1 – 120 немецких фраз;
4. ОВ 2.2 – 300 гласных фонем, сформированных из ОВ 2.1.

Реконструкция аттрактора

Для конструктивного решения задачи распознавания эмоций по речи необходимо количественно охарактеризовать речевой сигнал и выделить существенные параметры, отвечающие за эмоциональное состояние человека, т.е. необходимо подобрать соответствующий математический аппарат. Перспективным, по мнению авторов, в этом плане является аппарат нелинейной динамики, позволяющий реконструировать фазовый портрет аттрактора по временно́му ряду или по одной его координате. Для реконструкции аттрактора исследуемый временной ряд $x_n,...,x_{n-1}$ подвергается задержке координат [7]:

$$y_t = (x_t, x_{t+\tau},..., x_{t+(m-1)\tau}), \quad t = 0,...,s-1, \quad s = N-(m-1)\tau, \tag{1}$$

где N – общее число элементов (точек) временного ряда; τ – задержка по времени между элементами временного ряда (временной лаг); m – размерность вложения (размерность лагового пространства).

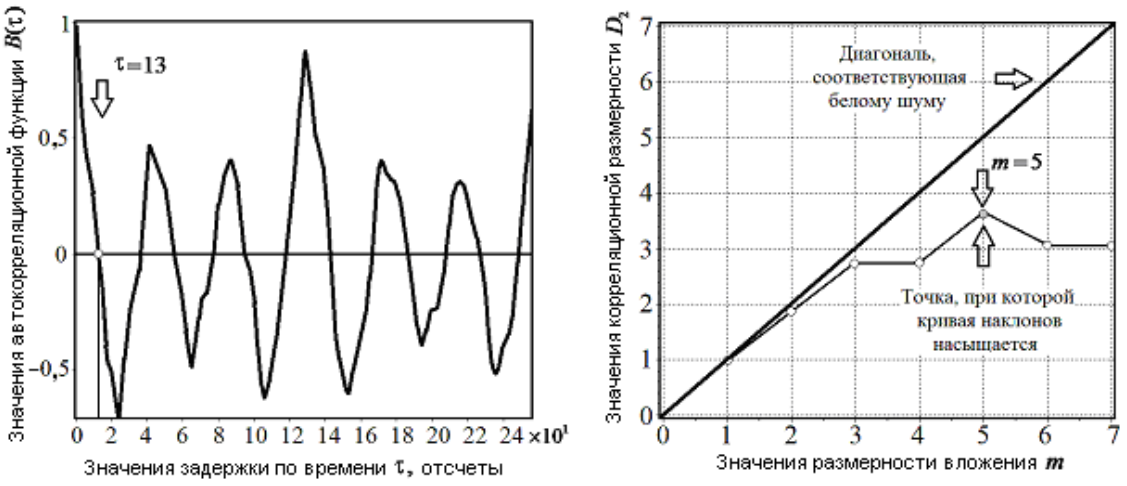
При выборе значения временной задержки τ используется идея о том, что если точки, образующие временной ряд, независимы друг от друга, то реконструированные вектора (1) несут в себе наибольшее количество информации об исследуемом ряде. По этой причине необходимо выбирать τ таким образом, чтобы корреляция между элементами временного ряда x_t и $x_{t+\tau}$ была по возможности минимальной. Такой выбор осуществляется при вычислении автокорреляционной функции $B(\tau) = 1/k \sum_{k=0}^{k-1}(x_n - \overline{x})\cdot(x_\tau - \overline{x}), \quad k = N-\tau$, где $\overline{x}$ – математическое ожидание. Временная задержка τ выбирается равной времени первого пересечения нуля автокорреляционной функцией [8] (рис. 1, а), значение задержки составляет 13 ($\tau = 13$). Величина размерности вложения m определяется с точки зрения достаточности (насыщения) посредством вычисления корреляционного интеграла $C(\varepsilon)$ и корреляционной размерности D_2 реконструкции аттрактора [8]. Корреляционный интеграл $C(\varepsilon)$, показывающий относительное число пар точек аттрактора x_i, x_j, находящихся на расстоянии не больше ε, определяется как

$$C(\varepsilon) = \lim_{M\to\infty} 1/M(M-1)\cdot\sum_{i,j=1}^{M}\theta(\varepsilon - r(x_i, x_j)), \quad i, j = 1,...,M, \tag{2}$$

$$D_2 = \lim_{\varepsilon\to 0} \log C(\varepsilon)/\log\varepsilon, \tag{3}$$

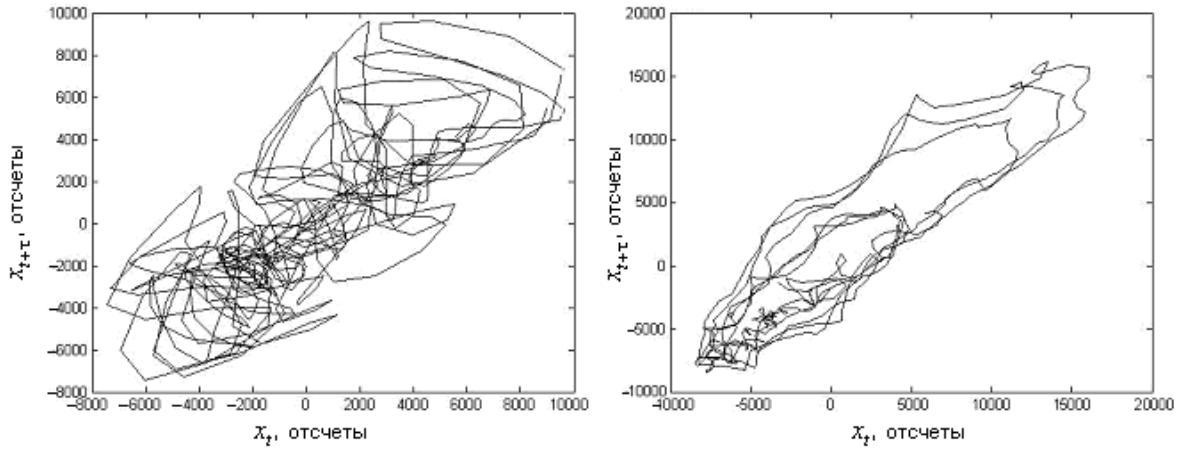
где M – число рассматриваемых состояний x_i (количество точек x_i на аттракторе); r – расстояние между точками аттрактора; $\theta(\alpha)$ – ступенчатая функция Хевисайда. После нахождения $C(\varepsilon)$ (2) и D_2 (3), строится зависимость корреляционной размерности D_2 от размерности вложения m (1), определяется точка, при которой кривая наклонов насыщается (рис. 1, б), корреляционная размерность аттрактора составляет 3,6 ($D_2 = 3,6$), она достигается при размерности вложения, равной 5 ($m = 5$).

Проведен нелинейный анализ фраз (ОВ 1.1, ОВ 2.1) и фонем (ОВ 1.2, ОВ 2.2) (рис. 2) на основе реконструкции аттрактора (1). Выявлено, что в большинстве случаев наблюдается взаимосвязь геометрии аттрактора с состоянием эмоционального возбуждения (объектам нейтрального состояния присуще более правильная форма, стремящаяся к эллипсообразной). Установлено, что эмоция радости по сравнению с нейтральным состоянием имеет меньшую траекторию разброса реконструкции, как для фраз, так и для фонем.



а б

Рис. 1. Выбор оптимальных значений реконструкции: автокорреляционная функция объекта ОВ 1.1 (а); зависимость значений D_2 от m (б)



а б

Рис. 2. Аттракторы фонемы «и» ОВ 1.2: радость (а); нейтральное состояние (б)

Предложен новый признак, определяемый по результатам реконструкции, который существенно снижает размерность описаний речевых образцов и позволяет осуществлять количественно сравнение аттракторов – усредненный максимальный вектор реконструкции аттрактора по четырем квадрантам $R_{\max}^{all}$. Вначале находится первый вектор реконструкции в первом квадранте $R_1^1 = \sqrt{x_i^2 + x_{i+\tau}^2}$, где x_i – значение временного ряда в i-й момент времени, τ – временная задержка. Далее вычисляются оставшиеся n векторов в первом квадранте, в результате получается множество значений векторов реконструкции $R^1 = \{R_1^1, R_2^1, ..., R_n^1\}$. Из множества R^1 выбирается максимальный вектор $R_{\max}^1$. Аналогично находятся максимальные вектора реконструкции аттрактора в других квадрантах $R_{\max}^2$, $R_{\max}^3$ и $R_{\max}^4$. Далее рассчитывается усредненный максимальный вектор реконструкции аттрактора по четырем квадрантам $R_{\max}^{all}$, который является новым количественным признаком для распознавания i-го речевого образца:

$$R_{\max}^{all}(i) = 0,25\sum_{j=1}^{4} R_{\max}^{j}(i), \quad i = 1,...,M, \tag{4}$$

где j – номер квадранта, i – номер речевого образца (предложение или фонема); $M = 18$ для ОВ 1.1; $M = 180$ для ОВ 1.2; $M = 120$ для ОВ 2.1; $M = 300$ для ОВ 2.2.

Количественная оценка реконструкций аттракторов на выборках речевых образцов разной длительности (таблица) выполнена с использованием следующих характеристик:

$$\overline{R}_{\max}^{all} = M^{-1} \cdot \sum_{i=1}^{M} R_{\max}^{all}(i), \quad \overline{R}_{\max}^{j} = M^{-1} \cdot \sum_{i=1}^{M} R_{\max}^{j}(i), \quad j = 1,...,4. \tag{5}$$

Установлено, что как на уровне фраз (ОВ 1.1, ОВ 2.1), так и на уровне фонем (ОВ 1.2, ОВ 2.2) эмоция радости по сравнению с нейтральным состоянием характеризуется меньшим значением $\overline{R}_{\max}^{all}$ (4),

(5). Следует отметить тот факт, что образцы русскоязычной части корпуса с эмоцией радости (на всех уровнях) имеют приблизительно в два раза меньшее значение признака $\overline{R}^{all}_{\max}$.

Объекты	Эмоциональное возбуждение	Выборка	Признаки, отсчеты				
			$\overline{R}^{1}_{\max}$	$\overline{R}^{2}_{\max}$	$\overline{R}^{3}_{\max}$	$\overline{R}^{4}_{\max}$	$\overline{R}^{all}_{\max}$
Фразы (предложения)	Радость	ОВ 1.1	19596	18786	16229	18561	18293
		ОВ 2.1	28257	34587	29716	39138	32925
	Нейтральное состояние	ОВ 1.1	37536	35547	31384	38358	35706
		ОВ 2.1	31671	33045	34846	40882	35111
Фонемы (звуки)	Радость	ОВ 1.2	13067	7969	9456	5361	8963
		ОВ 2.2	11098	10208	11800	11762	11217
	Нейтральное состояние	ОВ 1.2	28387	13795	18267	9194	17411
		ОВ 2.2	15590	11801	18777	14434	15151

Таблица. Усредненный максимальный вектор реконструкций аттракторов $\overline{R}^{all}_{\max}$

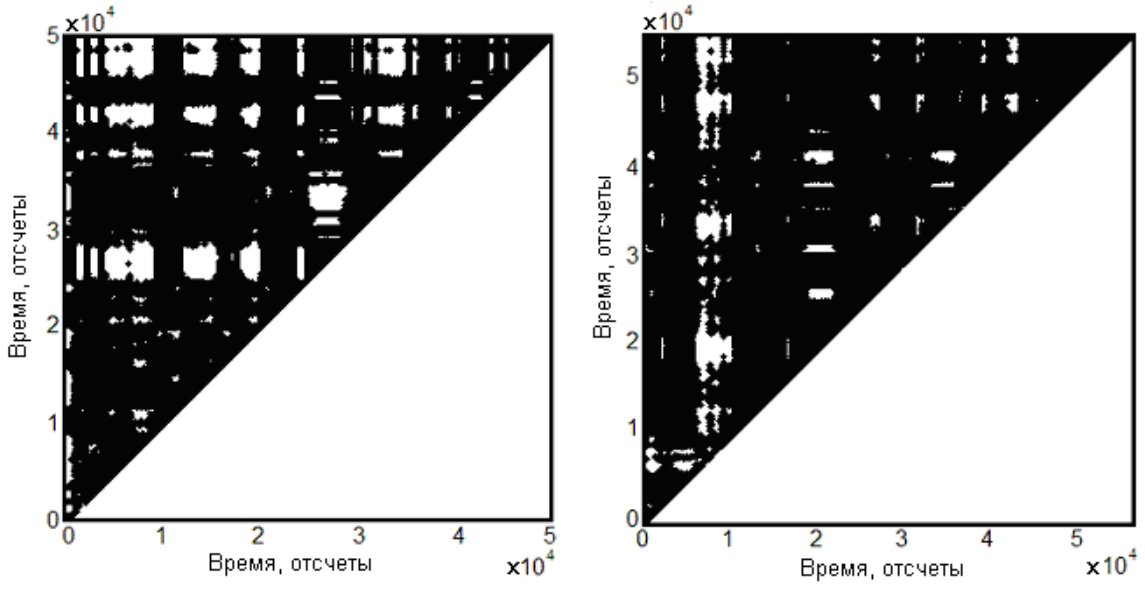
Рекуррентный график

В 1987 г. Экман и соавторы [9] разработали так называемые рекуррентные графики (диаграммы), позволяющие исследовать m-размерную траекторию лагового пространства (1) посредством двухмерного представления ее рекуррентности (повторяемости траекторий по происшествии некоторого времени в пространстве реконструкции аттрактора). Рекуррентный график представляется в виде двумерной или треугольной (так как обе стороны от главной диагонали под углом $\pi/4$ являются симметричными) матрицы размером $N \times N$, по обеим осям которой откладывается время. Матрица заполнена черными и белыми точками (единицами и нулями), где черные точки обозначают наличие рекуррентности, а белые – отсутствие [10]:

$$R_{ij} = \theta(\varepsilon_i - \|x_i - x_j\|), \quad i, j = 1,...,N, \tag{6}$$

где N – число рассматриваемых состояний x_i; ε_i – радиус выбранной окрестности (расстояние от центра окрестности x_i до ее границы); $\|\bullet\|$ – норма.

Если точка траектории реконструкции аттрактора в момент времени x_j попадает в выбранную окрестность другой точки в момент x_i, то такие точки считаются рекуррентными, вследствие чего на рекуррентном графике появляется точка черного цвета с координатами x_{ij}, соответствующая единице, и наоборот [7]. Радиус выбранной окрестности ε_i (6) выбирается не более 10% от максимального значения диаметра восстановленной реконструкции аттрактора [8]. На рис. 3 приведены примеры рекуррентных графиков объектов ОВ 1.1.



а б

Рис. 3. Рекуррентные графики фраз: радость (а); нейтральное состояние (б)

Визуально установлено, что для объектов, выражающих эмоцию радости (рис. 3, а), характерна более контрастная топология по сравнению с нейтральным состоянием (рис. 3, б). Эмоция радости характеризуется более резкими изменениями динамики временного ряда и нестационарностью, вследствие чего в структуре рекуррентного графика появляются характерно выраженные белые зоны, указывающие на нерегулярность процесса. Текстура эмоции радости характеризуется более выраженными скоплениями горизонтальных и вертикальных линий, повторяющихся с некоторой периодичностью.

Заключение

В среде MATLAB в виде m-файлов реализован программный модуль распознавания эмоции радости человека по речевому сигналу, основанный на использовании двух качественных (y_t, R_{ij}) и пяти количественных ($\overline{R}^1_{\max} - \overline{R}^4_{\max}$; $\overline{R}^{all}_{\max}$) признаков нелинейной динамики. При тестировании программного модуля на модельном корпусе эмоциональной речи точность распознавания, т.е. отнесения к одному из двух возможных классов (радость или нейтральное состояние), составила 93% для немецкоязычной и 95% для русскоязычной частей корпуса. Для сравнения отметим, что при распознавании образцов «нейтральной» и «агрессивной» речи из базы Emo-DB точность распознавания 96% получена при использовании 4 признаков, а 98% – при использовании 384 признаков [1]. В работе [2] классификатор, построенный для этой же базы Emo-DB, решал задачу разделения двух классов образцов речи (нормальное состояние и отклонение от него, возникающее у человека, испытывающего различные эмоции). Точность классификации составила 97 % при использовании 211 признаков и 87 % – при 15 признаках. Предлагаемый набор параметров аппарата нелинейной динамики после соответствующей адаптации будет использоваться для формирования динамической модели, отображающей взаимосвязь эмоционального состояния человека с характеристиками речевого сигнала.

Литература

1. Давыдов А.Г., Киселев В.В., Кочетков Д.С. Классификация эмоционального состояния диктора по голосу: проблемы и решения // Труды международной конференции «Диалог 2011». – М.: РГТУ, 2011. – С. 178–185.
2. Лукьяница А.А., Шишкин А.Г. Автоматическое определение изменений эмоционального состояния по речевому сигналу // Речевые технологии. – М.: Народное образование, 2009. – № 3. – С. 60–76.
3. Сидоров К.В., Филатова Н.Н. Анализ признаков эмоционально окрашенной речи // Вестник Тверского государственного технического университета. – Тверь: ТвГТУ, 2012. – Вып. 20. – С. 26–31.
4. Сидоров К.В., Филатова Н.Н., Калюжный М.В. Модельный русскоязычный корпус эмоциональной речи // Приоритетные направления развития науки и технологий: доклады XI всероссийской научн.-техн. конф. – Тула: Инновационные технологии, 2012. – С. 115–117.
5. Burkhardt F., Paeschke A., Rolfes M., Sendlmeier W., Weiss B. A Database of German Emotional Speech // Proc. Intern. Conf. Interspeech. – Lissabon, 2005 [Электронный ресурс]. – Режим доступа: http://pascal.kgw.tu-berlin.de/emodb/index-1280.html, свободный. Яз. англ. (дата обращения 10.07.2012).
6. Сидоров К.В., Филатова Н.Н. Алгоритм автоматической генерации речевых объектов // Сборник материалов I Международной научн.-практ. конф. «Технические науки – основа современной инновационной системы». – Ч. 1. – Йошкар-Ола, 2012. – С. 118–120.
7. Сидоров К.В. Диагностика эмоционального состояния диктора на основе рекуррентного анализа речевого сигнала // Междисциплинарные исследования в науке и образовании. – 2012. – № 1 Sp. – [Электронный ресурс]. – Режим доступа: http://www.es.rae.ru/mino/157-702, свободный. Яз. рус. (дата обращения 10.07.2012).
8. Горшков В.А., Касаткин С.А. Идентификация временных рядов авиационных событий методами и алгоритмами нелинейной динамики. – М.: Бланк Дизайн, 2008. – 208 с.
9. Eckmann J.P., Kamphorst S.O., Ruelle D. Recurrence Plots of Dynamical Systems // Europhys. Lett. 5. – 1987. – P. 973–977.
10. Киселев В.Б. Рекуррентный анализ – теория и практика // Научно-технический вестник СПбГУ ИТМО. – 2006. – № 29. – С. 118–127.

Сидоров Константин Владимирович – Тверской государственный технический университет, аспирант, bmisidorov@rambler.ru, bmisidorov@mail.ru

Филатова Наталья Николаевна – Тверской государственный технический университет, доктор технических наук, профессор, nfilatova99@mail.ru