



ИССЛЕДОВАНИЕ ЧАСТОТНО-ВРЕМЕННЫХ ХАРАКТЕРИСТИК РЕЧЕВЫХ СИГНАЛОВ И ОЦЕНОК ИХ ПРОИЗВОДНЫХ

Н. С. ТИТОВА

*Белгородский
государственный
университет*

e-mail: NTitova@bsu.edu.ru

В данной работе исследуются частотно-временные характеристики речевых сигналов, которые вычисляются с использованием коротких отрезков данных. К числу важнейших характеристик относятся доли энергии анализируемых отрезков сигнала и их производных, попадающие в заранее выбранные частотные интервалы.

Ключевые слова: оценка производной, частотное представление, доля энергии, устойчивость вычислений.

Для сжатия речевых сигналов необходимо осуществить некоторые преобразования (перекодирование), из которых не все являются обратимыми, в том смысле, что допускают полное восстановление исходных представлений (данных). Поэтому, одной из проблем сжатия является сохранение высокого качества воспроизведения восстанавливаемых звуков речи.

Целесообразно выделить два основных аспекта сжатия речевых данных: сжатие за счет кодирования пауз в речевых сообщениях, на основе фиксирования их начал и окончаний (что и хранится в окончателных данных); уменьшение объёмов битовых представлений данных о произнесённых звуках речи на основе соответствующих их преобразований. В последнем случае чаще всего речь идёт об использовании свойства сосредоточенности энергий отрезков звуков речи в малой доле частотной полосы. Именно это позволяет пренебречь мало энергетическими частотными интервалами.

Исследование свойств производных речевых сигналов позволяет описать (построить модели) звуков русской речи с точки зрения динамических характеристик, которые имеют важное значение для их восприятия механической системой человеческого слуха. В частности первая производная определяет скорость нарастания звуковых сигналов, а вторая – соответствующее ускорение. Производные более высоких порядков не имеют явной физической трактовки. Однако их значения могут служить важными признаками при сегментации речевых сигналов на звуки речи и шумы в паузах [1].

В частности для задач сжатия может оказаться важным распознавание и отдельных звуков речи, так как это позволит принимать решение о допустимой степени искажений, например, при разделении вокализованных, невокализованных и назальных звуков речи.

Соответствующие разным звукам речи отрезки речевых сигналов имеют различные длительности, а также отличаются характером распределения энергий вдоль частотной оси. Поэтому с точки зрения сжатия речевых данных их адекватное описание (модели) на основе оценок производных должно основываться на анализе частотно – временных характеристик, которые вычисляются с использованием коротких отрезков данных.

К числу важнейших характеристик относятся доли энергий анализируемых отрезков сигналов и их производных, попадающие в заранее выбранные частотные интервалы, совокупность которых покрывает всю частотную ось без перекрытий.

Значимость этой характеристики для сопоставлений поведений отрезков сигнала с поведением производных определяется в частности тем, что у производных должны усиливаться составляющие в высокочастотной области, тогда как в области низких частот эти составляющие наоборот должны ослабляться. В самом деле, пусть для непрерывного сигнала справедливо частотное представление



$$u(t) = \int_{-\Omega}^{\Omega} U(v) \exp(jvt) dv / 2\pi, \quad (1)$$

где предполагается, что область определения трансформанты Фурье является конечной, хотя и неопределённо большой ширины, что позволяет при дискретизации избежать искажений за счет наложения спектров из разных частотных областей (в соответствии с формулой Найквиста).

Ограниченность области определения трансформанты Фурье гарантирует непрерывность, как самого сигнала, так и его производных любого порядка. Поэтому на основе соотношения (1) нетрудно получить частотные представления для производных

$$u^{(k)}(t) = d^k u(t) / dt^k = \int_{-\Omega}^{\Omega} U(v)(jv)^k \exp(jvt) dv / 2\pi. \quad (2)$$

Отсюда следует, что трансформанта Фурье соответствующей производной

$$U_k(v) = \int_{-\infty}^{\infty} u^{(k)}(t) \exp(-jvt) dt \quad (3)$$

следующим образом связана с трансформантой Фурье исходного сигнала

$$U_k(v) = (-jv)^k U(v), \quad (4)$$

и, следовательно, должно выполняться тождество

$$|U_k(v)|^2 = v^{2k} |U(v)|^2. \quad (5)$$

Тогда доли энергий производных в заданном частотном интервале определяются представлениями

$$P_V^{(k)} = \int_{v \in V} |U_k(v)|^2 dv / 2\pi = \int_{v \in V} v^{2k} |U(v)|^2 dv / 2\pi. \quad (6)$$

Здесь частоты не предполагаются нормированными (умножение на шаг дискретизации).

Соотношение (6) иллюстрирует то, каким образом преобразуются распределения долей энергий производных по сравнению с исходным непрерывным сигналом.

Если условия Найквиста при дискретизации выполняются, то сохраняется и соотношения вида (6) между вычисляемыми долями энергий. При этом нетрудно получить пригодную для вычислений формулу

$$P^{(k)} = \vec{u}^{(k)T} A \vec{u}^{(k)} = \vec{f}^{(k-1)T} A \vec{f}^{(k-1)}. \quad (7)$$

Для корректных сопоставлений долей энергий следует осуществлять нормирование, вычисляя отношения

$$s^{(k)} = P^{(k)} / \|\vec{u}^{(k)}\|^2, \quad k = 0, \dots, 3, \quad (8)$$

где $\|\vec{u}^{(k)}\|^2$ – квадрат евклидовой нормы (энергия) отрезка анализа (где вычисляются оценки производных)

$$\|\vec{u}^{(k)}\|^2 = \sum_{i=1}^M (u_i^{(k)})^2. \quad (9)$$

Компоненты векторов здесь представляют собой отсчёты производных, оценки которых вычисляются в [1,3]. Отметим, что имеются в виду производные, вычисляемые, в том числе и без предварительной фильтрации при выделении формантной структуры.

По определению производные функций с неизменяющимися значениями будут равны нулю. Как, очевидно, следует из соотношения $v_i = u_i - u_{i-1}, i = 1, \dots, N$, оценка производной будет равна нулю, если одинаковы отсчёты сигнала на анализируемом отрезке. Ясно, что отсчёты анализируемого отрезка всегда можно преобразовать к виду (с учётом формантной структуры)

$$u_k = z_k + m_u, \quad k = 1, \dots, N, \quad (10)$$

где m_{ur} – среднее значение

$$m_u = \sum_{i=1}^N u_i / N \cdot \tag{11}$$

Поэтому компоненты вектора из $v_i = u_i - u_{i-1}, i = 1, \dots, N$, будут равны разностям отклонений от среднего

$$v_i = z_i - z_{i-1}, i = 1, \dots, N. \tag{12}$$

Таким образом, при сопоставлении распределений энергий у отрезка сигнала и соответствующих оценок производных может оказаться целесообразным вычитать из исходных данных их среднее значения, что возможно будет более корректным и повысит достоверность и обоснованность получаемых выводов.

На основе соотношений вида (7) выясняется насколько и в какую сторону в зависимости от звуков речи доли энергий производных изменяются по сравнению с долями энергий исходных речевых сигналов. При этом существенное значение имеет анализ влияния на эти характеристики интервалов времени развития звуков: начало звучания; середина; окончание звука.

Для апробации метода проводились вычислительные эксперименты. На первом этапе создается необходимый запас эмпирических данных в виде записей речевых сигналов, которые соответствуют всем звукам русской речи. Для сегментации записи по звукам речи на слух определяются начало и концы отрезков речевых сигналов, которые соответствуют только одному звуку русской речи. Именно эти отрезки затем и подвергаются анализу.

Для иллюстрации выводов некоторые наиболее характерные результаты экспериментов представляются в виде графиков, отображающих поведение сопоставляемых характеристик.

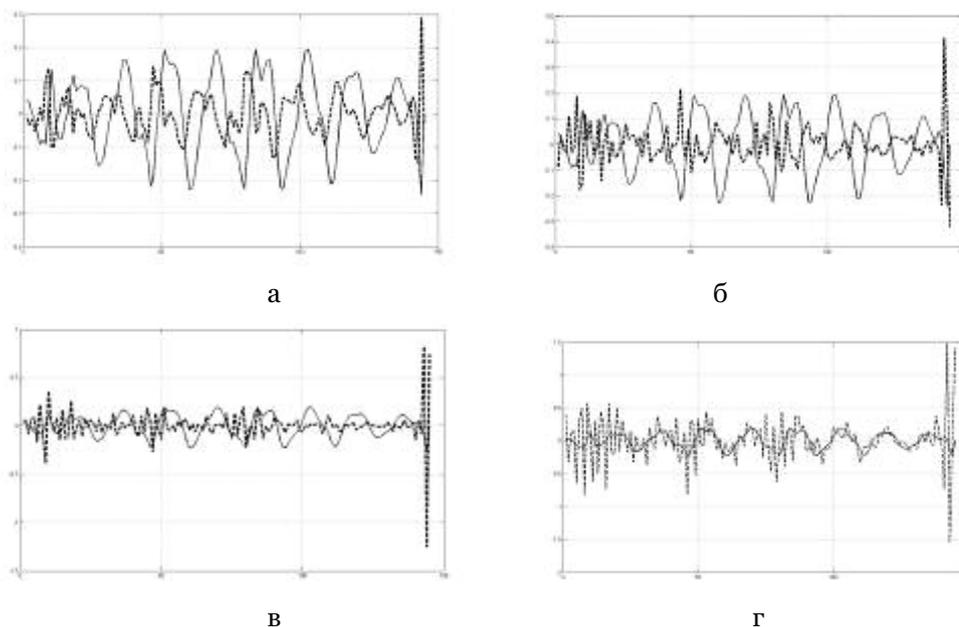


Рис. 1. Сигнал соответствующий звуку «о» (сплошная линия) и оценка его производных (штриховая линия): а) первая производная; б) вторая производная; в) третья производная; г) четвертая производная

На рис. 1 приводятся графики, иллюстрирующие поведение во времени оценок первой, второй, третьей и четвёртой производных речевых сигнала, порождаемых звуками русской речи. При этом в виду увеличения размахов колебаний оценок

производных при увеличении их порядка масштабы по осям ординат соответствующим образом изменяются.

Сравнительный анализ этих графиков позволяет сделать определённые выводы о различиях в поведении сигналов и их производных.

Для всех звуков речи, которые порождают сигналы с высокой концентрацией энергии в малой доле частотной полосы (до 35%) первые производные имеют меньший размах по сравнению с самим сигналом и только для производных высших порядков наблюдается превышение амплитуд колебаний над амплитудой колебания сигналов. При этом в колебаниях производных проявляются более быстрые изменения амплитуд. Это нетрудно объяснить тем, что высокочастотные компоненты сигналов при вычислении производных усиливаются, тогда как низкочастотные должны ослабляться вплоть до удаления постоянных составляющих. Особенно заметны эти проявления на графиках производных четвёртого порядка.

Таким образом, мощные низкочастотные составляющие речевых сигналов фильтруются при вычислениях производных.

Для речевых сигналов, порождаемых шипящими звуками речи, характерно наличие высокочастотных компонент. Это проявляется в том, уже амплитуды первых производных начинают превосходить амплитуды колебаний самих сигналов.

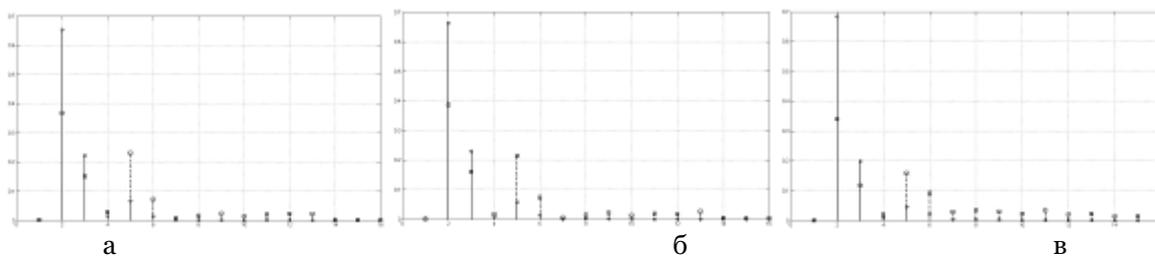


Рис. 2. Распределение долей энергии отрезков сигнала () и первой производной () по частотным интервалам ($R=16$, $N=128$) для звука «о»: а) начало звука; б) середина звука; в) окончание звука

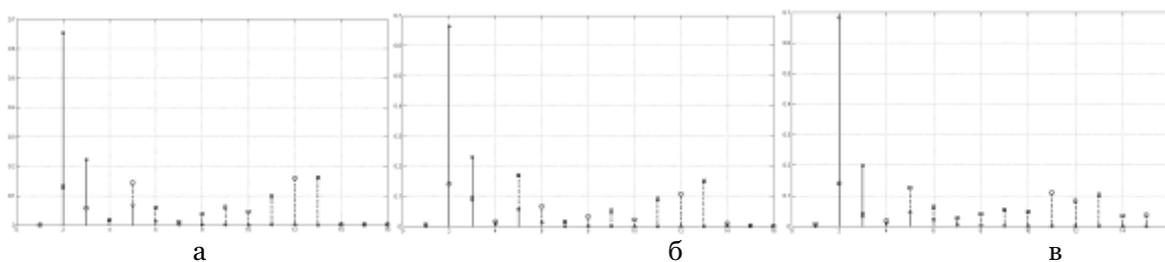


Рис. 3. Распределение долей энергии отрезков сигнала () и второй производной () по частотным интервалам ($R=16$, $N=128$) для звука «о»: а) начало звука; б) середина звука; в) окончание звука

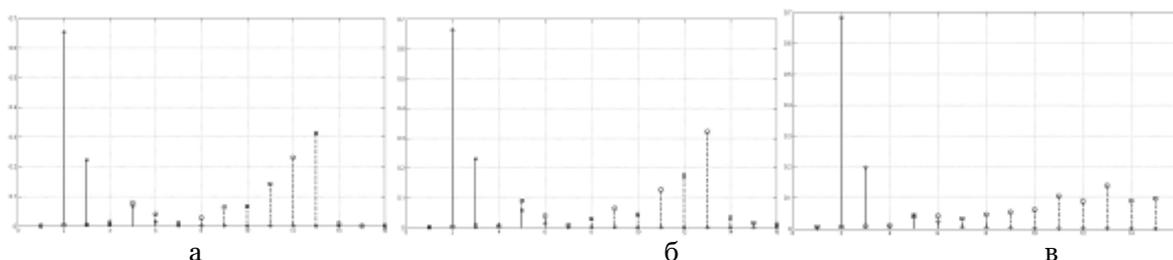


Рис. 4. Распределение долей энергии отрезков сигнала () и третьей производной () по частотным интервалам ($R=16$, $N=128$) для звука «о»: а) начало звука; б) середина звука; в) окончание звука

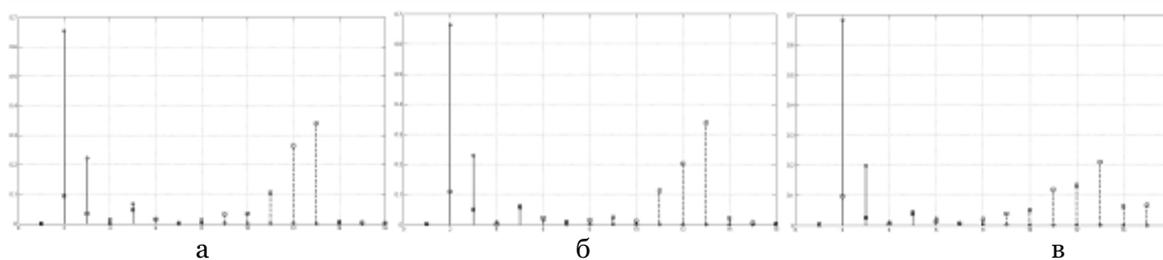


Рис. 5. Распределение долей энергии отрезков сигнала (Ψ) и четвертой производной (Φ) по частотным интервалам ($R=16$, $N=128$) для звука «о»: а) начало звука; б) середина звука; в) окончание звука

На рис. 2-5 приведены графики распределений энергий порождаемых звуками речи речевых сигналов и их производных. Они отражают динамику в поведении этих характеристик с точки зрения проявлений квазипериодических составляющих в виде концентрации энергий в узких частотных интервалах. Существенное значение имеет степень изменений концентраций энергий с ростом порядка производных, и в частности перераспределение её в пользу высокочастотных составляющих.

Свойства речевых сигналов до определённой степени зависят от интервала времени их регистрации: начало звука, середина либо окончание. Ясно, что именно свойства сигналов в начале и конце звуков играют основную роль в их сегментации, так как при этом следует учитывать зафиксированные различия, чтобы обоснованно принять решение.

С другой стороны, именно соответствующие середине звука отрезки сигналов наиболее характерно отражают различия в поведении сигналов и их производных. Поэтому на указанных рисунках сопоставляются распределения энергий сигналов и производных в течение трёх стадий развития звуков речи.

Литература

1. Жилияков Е.Г. Вариационные методы анализа и построения функций по эмпирическим данным: моногр. / Е.Г. Жилияков. – Белгород: Изд-во БелГУ, 2007г. – 160 с.
2. Жилияков, Е.Г. Вариационный метод оценивания производных и интерполяции сигналов по эмпирическим данным [Текст] / Е.Г. Жилияков, Т.Н. Созонова, И.Ю. Мисливец // Вестник Воронежского государственного университета, Серия: Системный анализ и информационные технологии. – Воронеж, 2006. – выпуск 2. – С.70-73.
3. Титова Н.С. О вычислении оценок производных высшего порядка по эмпирическим данным [Текст] / Титова Н.С. – Научные ведомости БелГУ, №9 (64), 2009, Выпуск 11/1.
4. Ланцош, К. Практические методы прикладного анализа [Текст] : справ. рук. / К. Ланцош; пер. с англ. М. З. Кайнера. – М. : Физматгиз, 1961. – 524 с.
5. Хургин, Я. И. Фinitные функции в физике и технике [Текст] / Я. И. Хургин, В. П. Яковлев. – М. : Наука, 1971. – 408 с. : ил.
6. Вержбицкий, В.М. Численные методы [Текст] / В.М. Вержбицкий. – М.: Высшая школа, 2000.

INVESTIGATION OF FREQUENCY-TEMPORAL CHARACTERISTICS OF SPEECH SIGNALS AND ESTIMATES OF THEIR DERIVATIVES

N. S. TITOVA

Belgorod State University

e-mail:

NTitova@bsu.edu.ru

In this paper we study the frequency-temporal characteristics of speech signals, which are calculated by using short segments of data. Among the most important characteristics are the proportion of energy of the analyzed segments of the signal and their derivatives, which fall in the pre-selected frequency range.

Key words: derivative estimation, frequency representation, the share of energy, the stability calculations.