

УДК [621.391 + 517.443] : 004.934

Анализ свойств самоподобия звукового сигнала на основе разностных и суммирующих алгоритмов дробного порядка

А.А. Жарких¹, И.А. Павлов²

¹ Судоводительский факультет МА МГТУ, кафедра радиотехники и радиотелекоммуникационных систем

² Политехнический факультет МГТУ, кафедра высшей математики и программного обеспечения ЭВМ

Аннотация. В работе исследуется самоподобие звукового сигнала на основе разностных и суммирующих алгоритмов дробного порядка. В алгоритмах использованы скользящие суммы и скользящие разности, описывающие низкочастотную и высокочастотную фильтрацию исходного звукового сигнала соответственно. Частотные коэффициенты передачи соответствующих фильтров подвергаются возведению в дробную степень. Представлены некоторые результаты компьютерного моделирования преобразований.

Abstract. In this paper audio signal self-similarity based on differential and integrating algorithms of fraction order has been researched. In algorithms the moving sums and the moving differences describing low-frequency and high-frequency filtering of the input audio signal accordingly have been used. The frequency transfer constants of appropriate filters have been raised to the fractional order. Some results of computer simulation of transformations have been presented.

Ключевые слова: самоподобие звукового сигнала, распознавание звукового сигнала, распознавание речи, признаки речевого сигнала, скользящие суммы и скользящие разности, низкочастотная и высокочастотная фильтрация, алгоритмы дробного порядка, преобразование спектра, дискретное преобразование Фурье

Keywords: audio signal self-similarity, audio signal recognition, speech recognition, speech features, moving sums and moving differences, low-pass and high-pass filtering, algorithms of fractional order, transformation of spectrum, discrete Fourier transformation

1. Введение

Звуковые сигналы обладают свойством самоподобия. Интуитивно это свойство можно понимать следующим образом. При некотором разделении спектра звукового сигнала на низкочастотную и высокочастотную составляющие они воспринимаются на слух с очень высокой степенью сходства с исходным сигналом, но различной громкости. Низкочастотная область звучит существенно громче, чем высокочастотная. Известно, что низкочастотный фильтр реализуется процедурой интегрирования, а высокочастотный – процедурой дифференцирования. Если пренебречь ослаблением высокочастотной составляющей, то можно считать, что звуковой сигнал квазиинвариантен относительно операций интегрирования и дифференцирования. Эта квазиинвариантность и отражает самоподобие сигнала.

Цель работы – исследование самоподобия звукового сигнала на основе разностных и суммирующих алгоритмов дробного порядка.

В работе (Жарких и др., 2006) приведены теоретические сведения, необходимые для реализации разностных и суммирующих алгоритмов дробного порядка. В работе (Шахтарин, 2002) представлено дифференцирование и интегрирование дробного порядка применительно к аналоговому сигналу. В нашей работе рассматривается дискретный аналог таких операций (Жарких и др., 2006). В работах (Павлов, Жарких, 2010; Жарких, Павлов, 2010) изложены результаты первых экспериментов по изучению самоподобия такими методами.

2. Разностные и суммирующие преобразования цифрового сигнала

Разностное преобразование определяется следующим образом:

$$y_{sub}(k) = x(k) - x(k - S), \quad (1)$$

где $x(k)$ – исходный сигнал, $y_{sub}(k)$ – преобразованный сигнал, k – номер отсчета, $k = 0, \dots, N-1$, N – число отсчетов, S – величина задержки.

Соответственно, спектры исходного сигнала и сигнала, полученного в результате разностного преобразования, связаны следующим образом:

$$Y_{sub}(m) = X(m) \cdot (1 - W_N^{-mS}), \quad (2)$$

где $X(m)$ – дискретный спектр исходного сигнала, $Y_{sub}(m)$ – дискретный спектр преобразованного сигнала, m – номер спектрального отсчета, $W_N = \exp(j \cdot 2\pi/N)$.

Суммирующее преобразование определяется формулой:

$$y_{sum}(k) = \frac{1}{S+1} \sum_{i=0}^S x(k+i), \quad (3)$$

где $y_{sum}(k)$ – преобразованный сигнал, $S+1$ – число усредняемых отсчетов, остальные параметры такие же, как в формуле (1).

Соответственно, спектр исходного сигнала и спектр сигнала, полученного в результате суммирующего преобразования, связаны следующим образом:

$$Y_{sum}(m) = \frac{1}{S+1} X(m) \frac{W_N^{-m(S+1)} - 1}{W_N^{-m} - 1}, \quad (4)$$

где $Y_{sum}(m)$ – дискретный спектр преобразованного сигнала, остальные параметры такие же, как в формуле (3).

Фактически выше представлены линейные стационарные фильтры, которые в частотной области могут быть описаны формулой:

$$Y(m) = K(m) \cdot X(m). \quad (5)$$

Частотный коэффициент передачи фильтра после нелинейного преобразования – возведения в дробную степень $1/n$ определяется выражением:

$$\tilde{K}(m) = |K(m)|^{1/n} \cdot e^{j(\varphi(m)/n)} \cdot e^{j(2\pi k/n)}, \quad k = 0, \dots, n-1, \quad (6)$$

где k – индекс, обозначающий номер ветви извлечения корня.

Многозначность этого преобразования приводит к многочисленным сценариям формирования разностных и суммарных сигналов дробного порядка, что проявляется в выходном звуковом сигнале при изменении его фазовых параметров.

3. Результаты моделирования

Было проведено два вида компьютерного эксперимента для анализа самоподобия звукового сигнала. В обоих экспериментах степень сходства двух сигналов оценивалась на основе прослушивания и сравнения графиков сигналов во временной и частотной областях.

В компьютерном эксперименте первого вида оценивалась степень сходства исходного сигнала с сигналом, подвергнутым разностному или суммирующему преобразованию дробной степени без изменения фазового спектра.

На рис. 1-3 представлены результаты компьютерного моделирования.

Использовался звуковой сигнал – фрагмент из оперы Бизе "Кармен". Число уровней квантования – 16, частота дискретизации – 22050 Гц. Параметры разностного и суммирующего алгоритмов: $S=1$, $q=1/2$.

Для временного представления сигнала единица измерения оси абсцисс – время в секундах, оси ординат – значение отсчета сигнала в диапазоне $[-1, 1]$; для амплитудного спектра сигнала по оси абсцисс отложена частота спектральных составляющих в герцах (Гц), а по оси ординат – значения амплитуд этих спектральных составляющих в децибелах (дБ).

На графиках во временной области видно, что после применения разностного преобразования уменьшилась амплитуда сигнала. На графиках амплитудных спектров можно увидеть, что разностное преобразование работает как высокочастотный фильтр, и низкие частоты частично гасятся, а суммирующее преобразование работает как низкочастотный фильтр, и частично гасится некоторая область высоких частот.

Прослушивание результатов преобразований позволяет сделать следующие выводы: сигнал после разностного преобразования степени $1/2$ звучит немного тише и более высокочастотно (звонко) относительно исходного сигнала; сигнал после суммирующего преобразования звучит с громкостью исходного, но более низкочастотно (глухо).

Прослушивание выходных сигналов в экспериментах первого вида показывает их существенное сходство с исходными сигналами. Это сходство лучше проявляется для записи речи и хуже – для музыки.

В компьютерном эксперименте второго вида оценивалось влияние фазовой последовательности на форму выходного звукового сигнала. Входной сигнал выбирался таким же, как в эксперименте первого вида.

Длина входных отсчетов сигнала дополнялась нулями до степени числа 2 ($N=2^n$). Выбиралось количество групп 2^p , $p = \{1, 2, \dots, n-1\}$. В каждой группе для всех отсчетов выбирался одинаковый номер корня. Для следующих друг за другом групп номер корня чередовался. Фазовая последовательность

задавалась в виде целых чисел $s(m)$, $0 \leq s(m) \leq 1$, $m = 1, \dots, (N/2 - 1)$. Комплексно-сопряженным элементам задавались разные знаки: $s(N-m) = -s(m)$. Значения $s(m)$ использовались в формуле (6) вместо k для соответствующего $K(m)$.

Амплитудный спектр для всех результатов преобразований соответствует рис. 2 (справа), поскольку изменение аргументов в комплексных числах не влияет на модули этих чисел.

Несколько характерных результатов, отражающих применение разностного алгоритма дробного порядка ($S=1$, $q=1/2$) с дополнительным введением фазовых последовательностей, представлены на рис. 4.

Оценки прослушивания результатов разностных преобразований ($S=1$, $q=1/2$) для различных значений параметра p приведены в таблице.

Таблица. Оценка влияния фазовой последовательности на форму выходного звукового сигнала

p	Оценка прослушивания
1,2	Нет значительных искажений сигнала, высокая степень подобия с исходным сигналом
3,4,5,6,7,8	Наличие короткого высокочастотного звука в конце сигнала
9,10,11,12	Фоновое повторяющееся эхо
13	Двукратное повторение искаженного отрезка сигнала
14	Перестановка во времени двух фрагментов сигнала

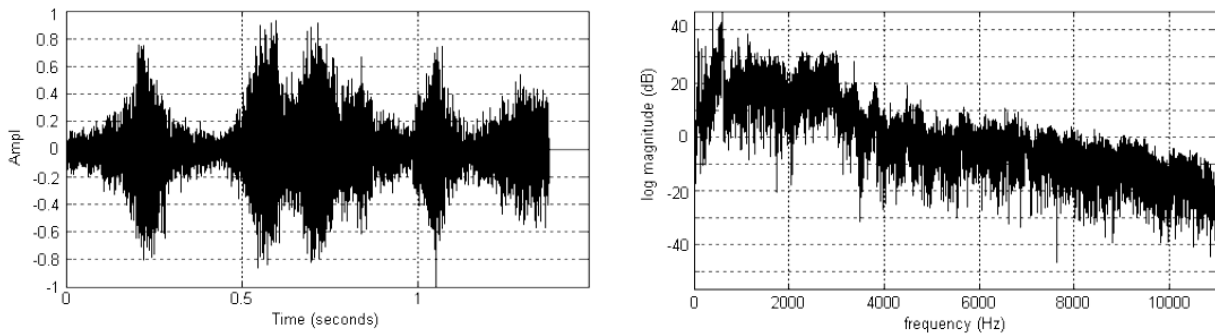


Рис. 1. Исходный звуковой сигнал

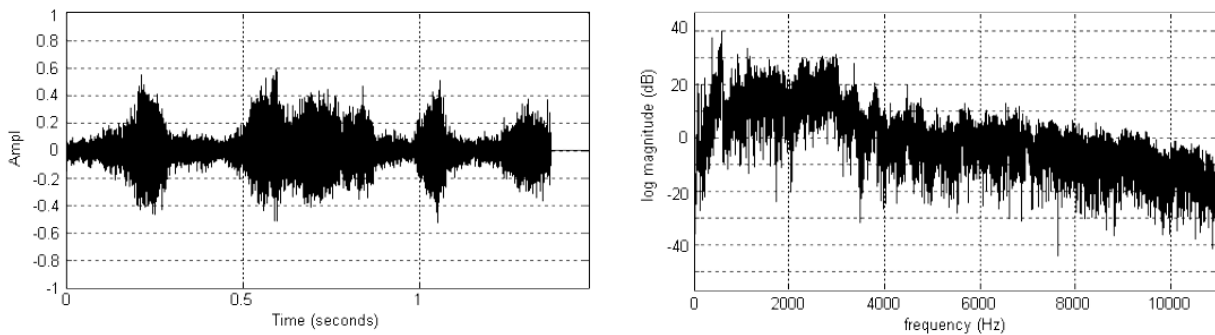


Рис. 2. Результат разностного преобразования при $S=1$, $q=1/2$

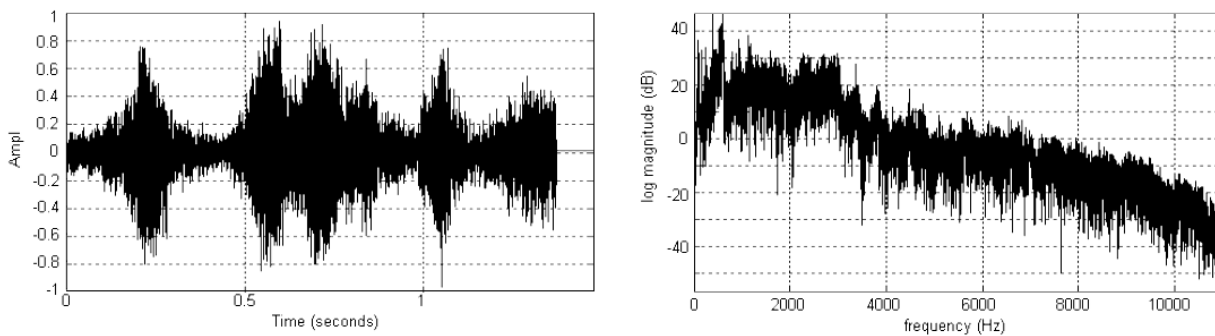


Рис. 3. Результат суммирующего преобразования при $S=1$, $q=1/2$

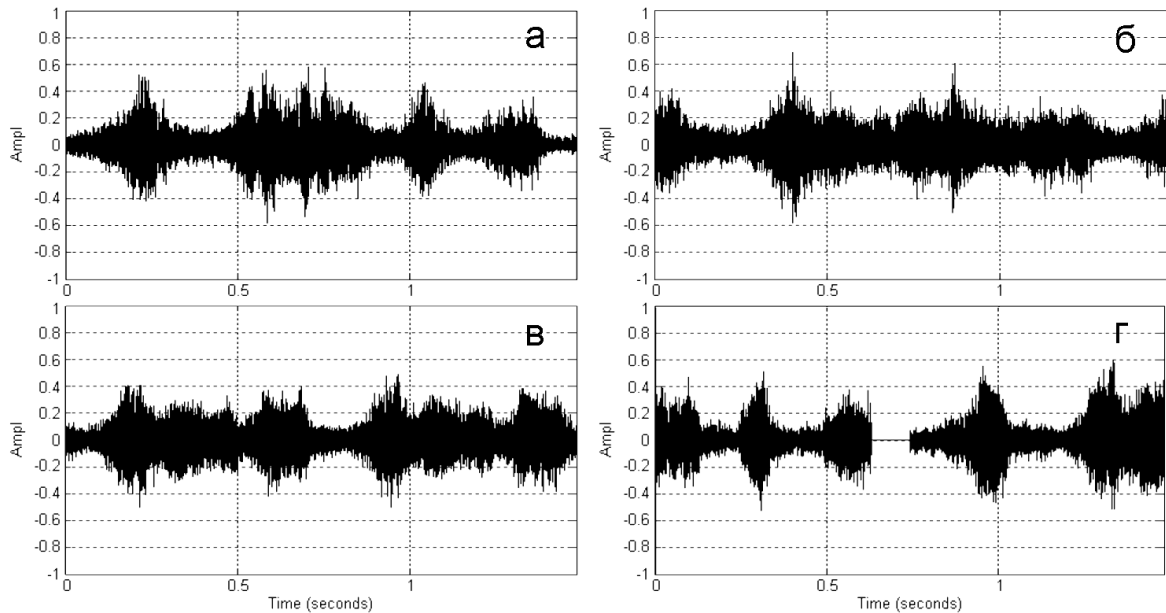


Рис. 4. Результаты разностного преобразования с введением фазовой последовательности для $S=1$, $q=1/2$; (а) $p=8$, (б) $p=12$, (в) $p=13$, (г) $p=14$

4. Заключение

Считается, что человеческий слух во многом подобен спектральному анализатору, т.е. ухо распознает амплитудный состав звуковых волн, а фаза практически не имеет значения (Калинцев, 1991; Чесебиев, 2008). В частности, системы распознавания речи обычно игнорируют фазу в речевом сигнале (O'Shaughnessy, 2003).

Результаты исследования позволяют сделать вывод, что разностные и суммирующие операции дробного порядка являются промежуточными между соответствующими операциями целого порядка.

Применение разностных и суммирующих операций как целого, так и дробного порядка показало высокую степень самоподобия речевых сигналов и более низкую степень самоподобия музыкальных произведений.

Предполагается использовать данный подход:

- для формирования дополнительных признаков речевого сигнала в задачах распознавания речи;
- для исследования механизмов речеобразования и речевосприятия;
- для уплотнения передачи сигналов путём введения дополнительной информации в виде фазовой последовательности.

Литература

- O'Shaughnessy D. Interacting with computers by voice: Automatic speech recognition and synthesis. *Proceedings of the IEEE*, v.91(9), p.1272-1305, 2003.
- Жарких А.А., Павлов И.А. Исследование свойств самоподобия речевого сигнала на основе разностных и суммирующих алгоритмов дробного порядка. *Материалы МНТК "Наука и образование – 2010"*, Мурманск, МГТУ, с.133-139, 2010.
- Жарких А.А., Степанов А.Н., Юрко А.С. Анализ самоподобия речевого сигнала на основе разностных и суммирующих алгоритмов дробного порядка. *Труды 61-й научной сессии, посвященной Дню радио*, М., НТОРЭС им. А.С.Попова, с.376-377, 2006.
- Калинцев Ю.К. Разборчивость речи в цифровых вокодерах. М., Радио и связь, 220 с., 1991.
- Павлов И.А., Жарких А.А. Реализация программного модуля для анализа самоподобия речевого сигнала на основе разностных и суммирующих алгоритмов дробного порядка. *Материалы межвузовского конкурса-конференции "Технологии Microsoft в теории и практике программирования" студентов, аспирантов и молодых ученых Северо-Запада, СПб., Изд-во Политехн. ун-та*, с.111-112, 2010.
- Чесебиев И.А. Компьютерное распознавание и порождение речи. М., Спорт и культура-2000, 128 с., 2008.
- Шахтарин Б.М. Случайные процессы в радиотехнике. М., Радио и связь, 568 с., 2002.