



АЛГОРИТМЫ ОБНАРУЖЕНИЯ ОСНОВНОГО ТОНА РЕЧЕВЫХ СИГНАЛОВ

Е.Г. ЖИЛЯКОВ
А.А. ФИРСОВА
Н.А. ЧЕКАНОВ

*Белгородский государственный
национальный исследовательский
университет*

e-mail: Zhilyakov@bsu.edu.ru

В статье представлено описание некоторых алгоритмов определения частоты основного тона речевых сигналов. Предложен новый алгоритм определения частоты основного тона звуков речи, основанный на двумерном методе обработки сигнала.

Ключевые слова: речевой сигнал, анализ речевого сигнала, частота основного тона, автокорреляционная функция, разностная функция, двумерная обработка речевого сигнала.

Одним из направлений развития современных информационно-телекоммуникационных систем является развитие технологий обработки речевых данных. К наиболее распространенным среди них относятся такие системы, как: идентификация по голосу, преобразование речи в текст, синтез по тексту, голосовое управление. Исследования особенностей распределения энергии звуков русской речи показали, что все звуки имеют свое особенное распределение энергии по частотным интервалам. Кроме того, распределение энергии зависит от местоположения звука, диктора, его эмоционального состояния и интонации. В свою очередь, среди характеристик речевых сигналов, соответствующих звукам русской речи, можно выделить те, которые незначительно изменяются на протяжении всего звука. Одним из таких параметров, который широко используется в системах распознавания и синтеза речевых сигналов, является частота основного тона. Частота основного тона – частота повторения колебаний голосовых связок при произнесении вокализованных звуков речи [1]. Колебания связок является одним из основных параметров источника голосового возбуждения речевого тракта. Они придают голосу звучание и характеризуют его высоту [2]. Значение частоты основного тона зависит от размеров и степени натяжения связок [3]. Среди оценок частоты основного тона принято выделять мгновенное значение частоты основного тона и среднее значение частоты основного тона. Проблема выделения частоты основного тона заключается в том, что эти значения могут изменяться. Значения частоты основного тона для разных дикторов находятся в диапазоне от 80 до 400 Гц. При этом для некоторых сигналов, соответствующих звукам русской речи, может более сильно проявляться частота обертонов (частоты кратные частоте основного тона, которые создают тембральный окрас диктора).

Все алгоритмы выделения частоты основного тона можно разделить на алгоритмы, основанные на: частотном анализе, временном анализе, учете корреляционных свойств речевых сигналов.

Одним из наиболее простых алгоритмов выделения частоты основного тона, который широко применялся в ряде приложений, является алгоритм, предложенный Голдом и усовершенствованный Рабинером и Голдом [3, 4]. Данный алгоритм основан на обработке сигнала во времени и учете его экстремумов. В основе метода лежит следующее положение: по речевому сигналу формируется несколько импульсных последовательностей, которые сохраняют периодичность входного сигнала и не содержат других его особенностей, бесполезных с точки зрения выделения основного тона.

На первом этапе сигнал сглаживается фильтром низких частот в диапазоне, соответствующем возможным значениям частоты основного тона (от 0 до 500 Гц). Далее определяются локальные минимумы и максимумы в сигнале. По их амплитуде и положению из отфильтрованного сигнала формируется несколько импульсных последовательностей. Каждая импульсная последовательность состоит из положительных импульсов,



возникающих в месте расположения максимума или минимума сигнала. Импульсные последовательности имеют следующий вид [3, 4]:

1. $m_1(n)$: импульс, равный по амплитуде значению локального максимума и формирующийся в месте расположения этого максимума;
2. $m_2(n)$: импульс, равный по амплитуде разности между максимумом и предшествующим минимумом и формирующийся в точке каждого максимума;
3. $m_3(n)$: импульс, равный по амплитуде разности между максимумом и предшествующим максимумом и возникающий в точке каждого максимума (если эта разность отрицательна, то импульс обращается в ноль);
4. $m_4(n)$: импульс, равный по амплитуде абсолютному значению локального минимума и формирующийся в месте расположения этого минимума;
5. $m_5(n)$: импульс, равный по амплитуде разности между максимумом и последующим минимумом и формирующийся в точке каждого минимума;
6. $m_6(n)$: импульс, равный по амплитуде разности между минимумом и предшествующим минимумом и возникающий в точке каждого минимума (если эта разность отрицательна, то импульс обращается в ноль).

В качестве оценок периода основного тона используются расстояния между пиками описанных импульсных последовательностей, превышающих заданный порог. Значение периода основного тона принимается равным наиболее часто встречающемуся значению оценок.

В рамках данной работы были проведены вычислительные эксперименты по оценке мгновенных и средних значений частоты основного тона для звуков русской речи одного диктора-женщины. Оценка мгновенных значений частоты основного тона проводилась на основе анализа отрезков одинаковой длины. Длительность отрезков анализа выбиралась равной $N=128$ отсчетов (16мс при частоте дискретизации 8к Гц). Выбор такой длины окна анализа позволяет учесть наименьшее возможное значение частоты основного тона. При этом такая длина достаточно мала, что позволит анализировать отрезки речевых сигналов, соответствующие одному звуку без захвата соседних звуков. Сдвиг окна анализа осуществлялся с шагом равным 1 отсчету (0,125 мс). Выбор такой величины шага позволяет проводить наиболее точный анализ изменения частоты основного тона. Решение о значении частоты основного тона принималось на основе определения наиболее часто встречающегося значения по всем импульсным последовательностям, и на основе среднего арифметического по всем импульсным последовательностям:

$$F_{OT}(n) = \frac{\sum_{k=1}^K f_{OT}(k, n)}{K}, \quad n=1, 2, \dots \quad (1)$$

где K – количество оценок частоты основного тона для всех импульсных последовательностей;

$f_{OT}(k, n)$ – k -ая оценка частоты основного тона для n -го окна анализа;

n – номер окна анализа.

Для принятия решения о среднем значении частоты основного тона для всего сигнала также использовалось два метода: оценка по наиболее часто встречающейся величине и оценка по среднему арифметическому с использованием выражения:

$$\bar{F}_{OT} = \frac{\sum_{n=1}^{N_{otr}} F_{OT}(n)}{N_{otr}}, \quad (2)$$

где N_{otr} – количество анализируемых отрезков сигнала;

$F_{OT}(n)$ – мгновенное значение частоты основного тона для n -го окна анализа;

n – номер окна анализа.

На рис. 1-2 представлены фрагмент речевого сигнала, соответствующего вокализованному звуку «а», и результаты оценки мгновенных значений частоты основного тона.

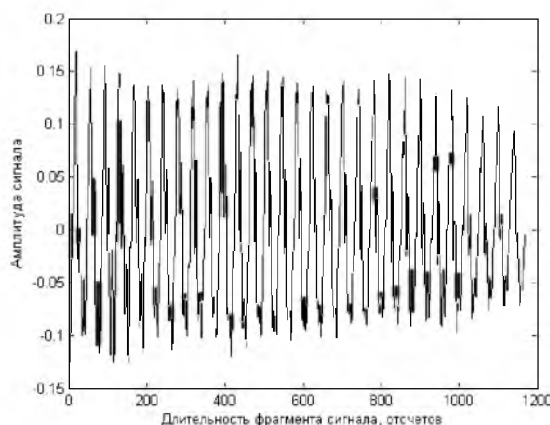


Рис. 1. Фрагмент сигнала, соответствующий ударному звуку «а» («атаман», $f_a=8\text{кГц}$)

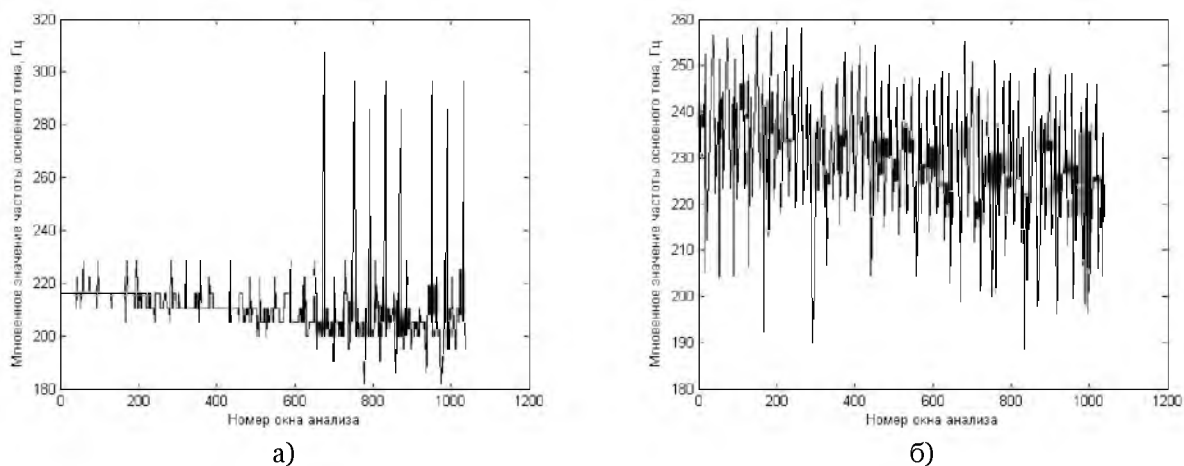


Рис. 2. График изменения мгновенных значений частоты основного тона для фрагмента сигнала, соответствующего звуку «а» («атаман», $f_a=8\text{кГц}$, $N=128$):

- а) принятие решение по большинству среди оценок импульсных последовательностей;
- б) принятие решение по среднему арифметическому оценок импульсных последовательностей

Анализ рисунков показывает, что величина частоты основного тона колеблется в диапазоне от 181 Гц до 307 Гц в случае принятия решения по большинству среди оценок импульсных последовательностей и от 188 Гц до 258 Гц в случае принятия решения по среднему арифметическому оценок импульсных последовательностей. Важно также отметить, что величина частоты основного тона для всего сигнала в первом случае составляет 216 Гц, а во втором 229 Гц. Оценка частоты основного тона при анализе отрезка не в автоматическом режиме показывает, что частота основного тона для данного фрагмента сигнала колеблется в диапазоне от 200 Гц до 228 Гц, среднее значение для всего сигнала составляет порядка 216 Гц. Таким образом, использование первого метода позволяет точнее определить среднее значение частоты основного тона.

Для оценки погрешности определения частоты основного тона использовались следующие параметры: наибольшее отклонение от среднего значения частоты основного тона, определенного не в автоматическом режиме и среднее квадратическое отклонение от среднего значения определенного не в автоматическом режиме. Наибольшее отклонение предлагается оценивать с использованием следующего выражения:

$$\sigma_{\max} = \max_{n=1 \dots N_{otr}} (|F_{OT}(n) - F_{OT}^{\text{не авт.}}|), \tag{3}$$

где N_{otr} – количество анализируемых отрезков сигнала;

$F_{OT}(n)$ – мгновенное значение частоты основного тона для n -го окна анализа;

n – номер окна анализа;

$F_{OT}^{\text{не авт.}}$ – частота основного тона, определенная не в автоматическом режиме.



Величину среднеквадратического отклонения предлагается оценить по формуле:

$$\sigma = \frac{\sum_{n=1}^{N_{otr}} (F_{OT}(n) - F_{OT}^{не\text{ авт.}})^2}{N_{otr}}, \quad (4)$$

где N_{otr} – количество анализируемых отрезков сигнала;

$F_{OT}(n)$ – мгновенное значение частоты основного тона для n -го окна анализа;

n – номер окна анализа;

$F_{OT}^{не\text{ авт.}}$ – частота основного тона, определенная не в автоматическом режиме.

Величина наибольшего отклонения для первого случая составляет 91 Гц, а для второго 42 Гц. В то время как величина среднеквадратического отклонения для первого случая составляет 14,69 Гц, а для второго – 18,74 Гц.

Результаты исследований показывают, что данный метод дает хорошие результаты на вокализованных сегментах речевого сигнала. Для невокализованных сегментов возникает значительный разброс в значениях оценок. Важно также отметить, что для некоторых дикторов в результате фильтрации возникают ситуации, когда в качестве частоты основного тона выбирается частота одного из обертонов. Реализация данного метода без использования предварительной фильтрации речевого сигнала приводит к возникновению большого числа экстремумов и, как следствие, возникновению ошибок при определении частоты основного тона.

Исследования особенностей изменения частоты основного тона для звуков русской речи показали, что для таких звуков как «г», «д», «з», «ж», «р» сложно определить частоту основного тона. Это связано с тем, что проявление частоты основного тона этих звуков существенно зависит от местоположения звука и его длительности. Для звуков «к», «п», «с», «т», «ф», «х», «ц», «ч», «ш», «щ» частота основного тона не проявляется. Это связано с природой этих звуков.

Исследование метода определения частоты основного тона для представленного алгоритма показало, что наибольшее среднеквадратическое отклонение от частоты основного тона при принятии решения по большинству среди оценок импульсных последовательностей наблюдается для звука «и» и составляет порядка 138 Гц. Для звуков «а», «й», «л», «н», «о», «у» данная величина находится в диапазоне от 100 Гц до 120 Гц. Для звуков «б», «в», «е», «ё», «м», «ы», «э», «ю», «я» наибольшее среднеквадратическое отклонение от частоты основного тона не превышает 50 Гц. Средняя величина среднеквадратического отклонения от частоты основного тона не превышает 51 Гц. В случае принятия решения о частоте основного тона по среднему арифметическому оценок импульсных последовательностей наибольшее среднеквадратическое отклонение от частоты основного тона наблюдается для звука «у» и составляет порядка 185 Гц. Для звуков «а», «и», «й», «л», «н», «о», «ю» данная величина находится в диапазоне от 100 Гц до 140 Гц. Для звуков «б», «в», «е», «ё», «м», «ы», «э», «я» наибольшее среднеквадратическое отклонение от частоты основного тона не превышает 74 Гц. Средняя величина среднеквадратического отклонения от частоты основного тона не превышает 55 Гц для всех перечисленных звуков русской речи, исключение составляет звук «ю», для которого среднее значение среднеквадратического отклонения составляет 139 Гц. Таким образом, при использовании алгоритма принятия решения по большинству среди оценок импульсных последовательностей средняя величина среднеквадратического отклонения меньше, чем в случае принятия решения по среднему арифметическому среди оценок импульсных последовательностей.

Другим способом определения частоты основного тона, нашедшим широкое применение в системах обработки речевых сигналов является автокорреляционный метод [3, 4, 5]. Суть метода состоит в том, что автокорреляционная функция отражает периодические свойства сигнала. Для любого периодического сигнала автокорреляционная функция достигает максимума в точках кратных периоду сигнала.

На первом этапе анализа аналогично описанному ранее методу необходимо осуществить сглаживание сигнала фильтром нижних частот с частотой среза 500 Гц. Для определения частоты основного тона предлагается использовать выражение вида:



$$R_n(p) = \sum_{m=0}^{N-1} [x(n+m) \cdot x(n+m+p)], \quad (5)$$

где $x(n)$ – отсчеты сигнала, соответствующего звуку русской речи,

N – длительность речевого сигнала,

p – порядок модели, характеризующий величину сдвига.

Порядок модели предлагается изменять в диапазоне от 20 до 100 отсчетов (для частоты дискретизации 8 кГц от 2,5 мс до 12,5 мс). Выбор таких параметров обусловлен ограниченностью значений частоты основного тона.

Анализ результатов исследования показывает, что наряду с пиком в точке значения периода основного тона автокорреляционная функция имеет побочные локальные максимумы, обусловленные затухающими колебаниями. Для снижения влияний этих колебаний в [3, 4] предлагается использовать методы центрального ограничения вида:

$$x_1(i) = \begin{cases} x(i), & \text{при } x(i) \notin (C \cdot x_{\min}, C \cdot x_{\max}) \\ 0, & \text{при } x(i) \in (C \cdot x_{\min}, C \cdot x_{\max}) \end{cases}, \quad (6)$$

$$x_2(i) = \begin{cases} 1, & \text{при } x(i) \notin (C \cdot x_{\min}, C \cdot x_{\max}) \\ 0, & \text{при } x(i) \in (C \cdot x_{\min}, C \cdot x_{\max}) \end{cases}, \quad (7)$$

где x_{\max} – максимальное значение амплитуды сигнала на интервале анализа,

x_{\min} – минимальное значение амплитуды сигнала на интервале анализа,

C – уровень ограничения (в [3, 4] предлагается использовать $C=0,68$).

Значение периода основного тона принимается равным точке максимума автокорреляционной функции. При этом важно отметить, что для вокализованных звуков амплитуда этого максимума составляет более 30% от значения автокорреляционной функции при порядке модели равном 0.

В рамках данной работы были проведены исследования оценки мгновенных и средних значений частоты основного тона при анализе сигнала без предварительной фильтрации, с применением сглаживания фильтром низких частот, а также при использовании методов центрального ограничения вида (6) и (7).

Исследования показали, что наибольшие значения среднеквадратического отклонения наблюдаются при использовании метода центрального ограничения вида (7). В остальных случаях среднеквадратическое отклонение примерно одинаковое. Исследования показали, что наибольшее значение среднеквадратического отклонения при отсутствии предварительной фильтрации наблюдается для звука «и» и составляет 178 Гц. Для таких звуков как «а», «в», «у», «л», «н» данная величина находится в диапазоне от 100 Гц до 128 Гц. Для звуков «б», «е», «ё», «й», «м», «о», «ы», «э», «ю», «я» величина среднеквадратического отклонения не превышает 93 Гц. При этом среднее значение величины среднеквадратического отклонения не превышает 35 Гц. Использование предварительной фильтрации и метода центрального ограничения вида (6) не позволяет достичь значительного выигрыша.

В работе [6] для выделения основного тона предлагается использовать разностную функцию вида:

$$S_i(\tau) = am / \left(\sum_{i=1}^{\tau} |x(t+i) - gain \cdot x(t+\tau+i)| + \delta \cdot am \right), \quad (8)$$

$$am = \sum_{i=1}^{\tau} |x(t+i)|, \quad (9)$$

$$gain = am / \sum_{i=1}^{\tau} |x(t+\tau+i)| \quad (10)$$

где $x(t)$ – отсчеты сигнала, соответствующего звуку русской речи,

τ – задержка сигнала,

δ – позволяет регулировать соотношение между максимумами разностной функции.

К достоинствам разностной функции вида (8) можно отнести меньшие вычислительные затраты по сравнению с автокорреляционной функцией.



В рамках данной работы были проведены исследования мгновенных и средних значений частоты основного тона для звуков русской речи. При этом задержка сигнала изменялась в диапазоне от 20 до 100 отсчетов (для частоты дискретизации 8 кГц от 2,5 мс до 12,5 мс). Реализация данного метода осуществлялась двумя способами: с использованием предварительной фильтрации и без.

Исследования показали, что наибольшее значение среднеквадратического отклонения при отсутствии предварительной фильтрации наблюдается для звука «и» и составляет 125 Гц. Для таких звуков как «а», «й», «у», «н» данная величина находится в диапазоне от 100 Гц до 109 Гц. Для звуков «б», «в», «е», «ё», «л», «м», «о», «ы», «э», «ю», «я» величина среднеквадратического отклонения не превышает 99 Гц. При этом среднее значение величины среднеквадратического отклонения не превышает 34 Гц. При использовании предварительной фильтрации наибольшее значение среднеквадратического отклонения достигается для звука «и» и составляет 128 Гц. Для звука «а» наибольшее значение среднеквадратического отклонения составляет 108 Гц. Для остальных звуков: «б», «в», «е», «ё», «й», «л», «м», «н», «о», «у», «ы», «э», «ю», «я» – наибольшее значение среднеквадратического отклонения не превышает 96 Гц. В этом случае среднее значение среднеквадратического отклонения не превышает 34 Гц.

В данной работе предлагается использование нового алгоритма выделения частоты основного тона. Суть метода состоит в представлении сигнала в виде изображения (двумерного массива). Такой подход направлен на поиск общих характеристик отрезков.

Сигнал представляется в виде двумерного массива размерности $M \times N$ [7]:

$$F(t, n) = x(t + n), \quad n = 1, 2, \dots, N, \quad t = 1, 2, \dots, M, \quad (11)$$

где $x(t)$ – отсчеты сигнала, соответствующего звуку русской речи,

N – длина окна анализа,

M – количество окон анализа.

Для оценки свойств сигнала рассчитывается матрица вида:

$$FF = \begin{cases} F \cdot F^T, & \text{если } M < N \\ F^T \cdot F, & \text{если } M \geq N \end{cases}, \quad (12)$$

Матрица FF является симметричной и неотрицательно определенной, а следовательно, обладает полной системой ортонормальных собственных векторов, соответствующих неотрицательным собственным числам:

$$Q \cdot L = FF \cdot Q, \quad (13)$$

$$Q = \{\vec{q}_1, \vec{q}_2, \dots, \vec{q}_N\}, \quad (14)$$

$$L = \text{diag}(\lambda_1, \dots, \lambda_N), \quad (15)$$

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N, \quad (16)$$

где Q – матрица собственных векторов матрицы FF ,

L – матрица собственных чисел,

q_i – собственный вектор,

λ_i – собственное число.

Строки матрицы собственных векторов отражают общие особенности элементов строк матрицы F . Через матрицу собственных векторов можно выразить матрицу, столбцы которой отражают общие признаки в каждой строке матрицы F :

$$U(:, i) = F \cdot \vec{q}_i / \sqrt{\lambda_i}, \quad i = 1, \dots, N \quad (17)$$

где F – матрица, состоящая из анализируемых отрезков сигнала,

q_i – собственный вектор,

λ_i – собственное число,

N – длина окна анализа.

Исследование матрицы U показало, что основные особенности речевых сигналов отражены в векторах, соответствующих первым двум собственным числам.

На рис. 3-4 представлены фрагмент сигнала, соответствующего звуку «а», и вектора матрицы U , соответствующие первым двум собственным числам.

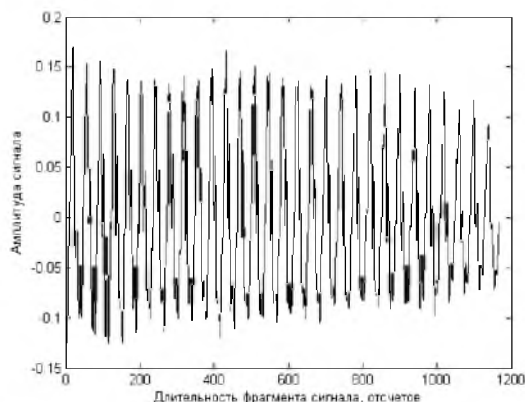


Рис. 3. Фрагмент сигнала, соответствующий ударному звуку «а» («атамАн», $f_d=8\text{кГц}$)

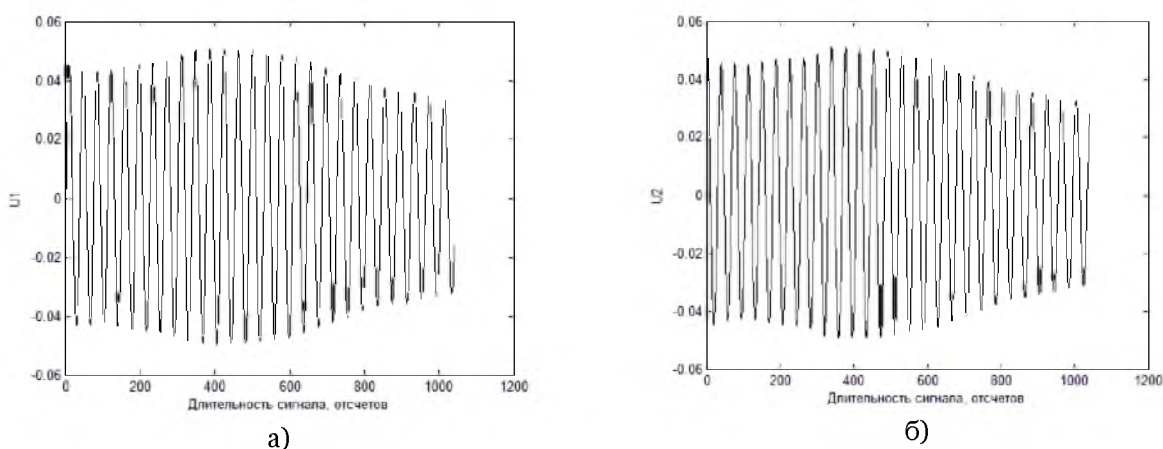


Рис. 4. График значений вектора матрицы U для фрагмента сигнала, соответствующего звуку «а» («атамАн», $f_d=8\text{кГц}$, $N=128$): а) для первого собственного числа; б) для второго собственного числа

Анализ рисунков показывает, что представленные вектора отражают колебание сигнала, имеющее наибольшую энергию. Для выделения огибающей предлагается использовать выражение вида:

$$Y(t) = \sqrt{U(t,1)^2 + U(t,2)^2}, \quad t=1, 2, \dots, M \quad (18)$$

где $U(t,1)$ – значения собственного вектора, соответствующего первому собственному числу; $U(t,2)$ – значения собственного вектора, соответствующего второму собственному числу.

На рис. 5 представлен результат оценки огибающей сигнала.

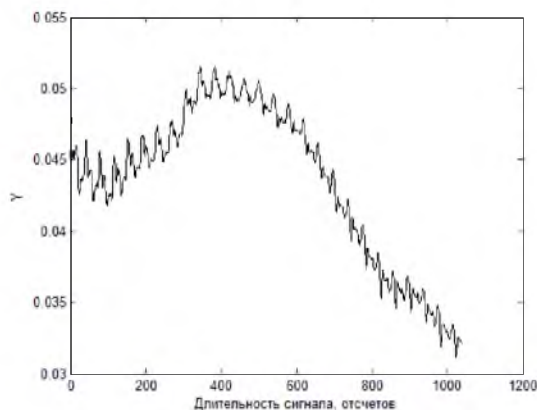


Рис. 5. График огибающей для векторов матрицы U , соответствующих первому и второму собственному числу (звук «а», «атамАн», $f_d=8\text{кГц}$, $N=128$)



Анализ рис. 4-5 показывает, что графики значений векторов матрицы U , соответствующих первому и второму собственным числам, позволяет выявить частоту основного тона анализируемого отрезка сигнала. Для определения частоты основного тона предлагается весь отрезок разбить на окна одинаковой длины, выбирая отрезки со сдвигом равным одному отсчету относительно начала окна анализа. Для каждого окна анализа определить распределение энергии по частотной оси. Частоту основного тона принять равной частоте с наибольшим значением энергии в диапазоне от 80 Гц до 400 Гц.

В рамках данной работе проводилась оценка частоты основного тона на основе анализа векторов матрицы U , соответствующих первому и второму собственному векторам, и на основе анализа вектора Y вида (18).

Исследования показали, что наибольшее отклонение от среднеквадратического отклонения наблюдается при использовании анализа вектора Y . При использовании анализа векторов матрицы U , соответствующих первому и второму собственным числам, наибольшее отклонение от величины среднеквадратического отклонения наблюдается для звука «а» и составляет 123 Гц. Для таких звуков как: «и», «й», «л», «н», «о», «у», «э» величина отклонения находится в диапазоне от 100 Гц до 112 Гц. А для звуков «б», «в», «е», «ё», «м», «ы», «ю», «я» – не превышает 79 Гц. Среднее значение среднеквадратического отклонения по всем звукам речи не превышает 33 Гц. Реализация алгоритма на основе анализ вектора Y приводит к большему количеству ошибок. Наибольшее отклонение наблюдается для звука «а» и составляет 141 Гц. Для звуков «м», «б», «е», «и», «о», «у», «ы», «н» эта величина находится в диапазоне от 100 Гц до 135 Гц. А для звуков «ю», «й», «я», «э», «л», «в» – не превышает 95 Гц. Среднее значение отклонения от величины среднеквадратического отклонения не превышает 75 Гц.

Таким образом, сравнение представленных алгоритмов показывает, что предлагаемый метод, основанный на двумерной обработке сигнала, позволяет более точно определить значения частоты основного тона речевых сигналов.

Основные проблемы определения частоты основного тона на вокализованных участках возникают в случае появления таких звуков как: «а», «и», «у». Ошибки возникают на участках начала и конца звуков, в случаях, когда энергия обертонов становится сопоставима с энергией частоты основного тона. При принятии решения о значении частоты основного тона диктора предлагается анализировать больший отрезок сигнала, соответствующего ударному гласному. Для принятия решения о мгновенном значении частоты основного тона предлагается использовать дополнительные алгоритмы, основанные на оценке нескольких возможных значений и принятии решения на основе наиболее вероятного из этих значений.

Исследования выполнены при поддержке гранта РНПВШ, госконтракт №8.2251.2011.

Список литературы

1. Лузин, Д.А. Разработка и исследование системы автоматического выделения основного тона речи : автореферат / Д.А. Лузин. – Ижевск, 2009.
2. Бабкин В.В. Помехоустойчивый выделитель основного тона речи / В.В. Бабкин // Труды 7-й Междунар. конф. и выставки Цифровая Обработка Сигналов и ее Применение (DSPA-2005) – Москва, 2005.
3. Рабинер, Л. Теория и применение цифровой обработки сигналов / Л. Рабинер, Б.Гоулд– М.: Мир, 1978. – 848с.
4. Рабинер, Л.Р. Цифровая обработка речевых сигналов / Л.Р. Рабинер, Р.В. Шафер – М.: Радио и связь, 1981. – 496с.
5. Шелухин, О.И. Цифровая обработка и передача речи [Текст] / О.И. Шелухин, Н.Ф.Лукьянцев; под ред. О.И. Шелухина. – М.: Радио и связь, 2000. – 456с.: ил.
6. Баронин, С.П. Автокорреляционный метод выделения основного тона речи. Пятьдесят лет спустя /С.П. Баронин// Речевые технологии, 2/2008, 2008, стр.3-12.
7. Жилияков Е.Г. Алгоритм сжатия речевых данных на основе двумерной обработки данных/Е.Г. Жилияков, А.В. Болдышев, Е.И. Прохоренко// Вопросы радиоэлектроники, серия ЭВТ, выпуск 1, 2012 – 27-33.



DETECTION ALGORITHM OF THE FUNDAMENTAL TONE SPEECH SIGNALS

E.G. ZHILYAKOV

A.A. FIRSOVA

N.A. CHEKANOV

*Belgorod National
Research University*

e-mail: Zhilyakov@bsu.edu.ru

This article describes some of the algorithms for the fundamental frequency of speech signals. A new algorithm for determining the fundamental frequency of speech sounds, based on the method of two-dimensional signal processing.

Key words: speech signal, speech signal analysis, fundamental frequency, the autocorrelation function, the difference function, two-dimensional processing of the speech signal.