УДК 621.391

ОЦЕНИВАНИЕ ПЕРИОДА ОСНОВНОГО ТОНА ЗВУКОВ РУССКОЙ РЕЧИ

Е.Г. ЖИЛЯКОВ
А.А. ФИРСОВА

Белгородский государственный национальный исследовательский университет

e-mail: zhilyakov@bsu.edu.ru

В статье представлен новый метод оценивания периода основного тона, основанный на использовании субполосного анализа и оценке корреляции составляющей каждой частотной полосы.

Ключевые слова: речевой сигнал, анализ речевого сигнала, частота основного тона, субполосный анализ.

Развитие информационных технологий направлено на обеспечение взаимодействия человека с техникой в наиболее удобной для человека форме. Наиболее популярными в этой области являются такие технологии как: распознавание речевых команд, преобразование речи в текст, распознавание и верификация дикторов. Реализация данных систем основана на анализе речевых данных с позиции выявления характеристик, позволяющих определить тип звука или же выделить особенности голоса диктора. Одной из таких характеристик является период основного тона. Период основного тона – величина обратная частоте основного тона, которая в свою очередь определяется частотой повторения возбуждающих воздействий гортани [1, 2]. Колебания связок является одним из основных параметров источника голосового возбуждения речевого тракта. Они придают голосу звучание и характеризуют его высоту [1, 2]. Значения частоты основного тона для разных дикторов находятся в диапазоне от 80 до 400 Гц. Значения частоты основного тона могут изменяться во времени, что определяет проблему выделения частоты основного тона.

Периодом основного тона принято считать интервал времени между двумя возбуждающими воздействиями. При этом речевой сигнал, взятый через период основного тона, почти повторяет свою форму. На рис. 1 представлен фрагмент речевого сигнала, соответствующего звуку «А», с указанием периодов основного тона.

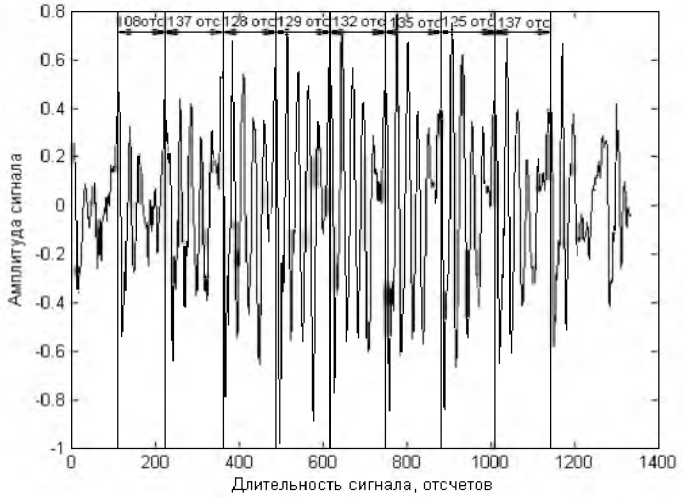


Рис. 1. Фрагмент сигнала, соответствующего звуку «А»

В настоящее время существует два основных подхода к определению частоты основного тона: на основе анализа спектров и корреляционном анализе.

Суть метода оценивания частоты основного тона заключается в определении значения частоты с максимальным значением энергии в диапазоне возможных значений частоты основного тона.

Основной недостаток спектрального оценивания заключается в следующем. Пусть последовательность отсчетов сигнала $(x_1,x_2,...)$ имеет периодический характер, так что

$$x_{i+kM}=x_i, \quad k=0,1,... \qquad (1)$$

Тогда соответствующая трансформанта Фурье (спектр):

$$X(\omega) = \sum_{i=1}^{LM} x_i e^{-j\omega(i-1)} \tag{2}$$

может быть представлена в виде:

$$X(\omega) = \sum_{k=1}^{L} e^{-j\omega M(k-1)} \sum_{i=1}^{M} x_i e^{-j(i-1)} .$$

Таким образом:

$$|X(\omega)|^2 = |X_p(\omega)|^2 \cdot \sin^2(LM\omega/2)/\sin^2(M\omega/2), \tag{3}$$

где $X_p(\omega)$ – спектр одного периода сигнала.

Легко понять, что первый сомножитель будет достигать максимального значения в следующих точках оси частот:

$$\omega_m = m2\pi/M, \quad m=1,2,..., \tag{4}$$

причем именно значение $2\pi/M$ соответствует частоте основного тона.

Однако влияние $|X_p(\omega)|^2$ может проявляться в том, что максимум правой части будет соответствовать другому значению m. Именно это не позволяет методически надежно определять период основного тона по спектру анализируемого отрезка сигнала.

В основе корреляционного метода определения периода основного тона используется характеристика:

$$\rho_{\tau,N} = \sum_{i=1}^{N} x_i x_{i+\tau} / \sqrt{\sum_{i=1}^{N} x_i^2 \sum_{k=1}^{N} x_{k+\tau}^2} , \tag{5}$$

которая является оценкой нормированного коэффициента корреляции.

В качестве значения периода принимается:

$$M = \arg\max \rho_{\tau,N} , \quad 0 \le \tau \le K , \tag{6}$$

где K определяется отношением частоты дискретизации к минимально возможной частоте основного тона.

Дополняющим к (6) условием является неравенство:

$$\rho_{M,N} \ge h \in (0.7 \div 1). \tag{7}$$

То есть максимальное значение характеристики (5) должно превышать некоторый порог, что отвечает условию почти периодического поведения отрезков сигнала на периоде.

Одним из недостатков такого подхода является присутствие искажающих шумов, что маскирует наличие периодичности в сигнале.

Кроме того, концентрация спектра $|X(\omega)|^2$ вблизи частоты, не совпадающей с $2\pi/M$, приводит к тому, что максимальное значение (5) будет достигаться при меньшем, чем длина интервала между возбуждающими гортань воздействиями.

Таким образом, необходимо использовать иные методы определения частоты основного тона, устойчивые как к воздействию шумов, так и к влиянию периодичности сигнала между двумя последовательными возбуждающими гортань воздействиями.

Представляется естественным ориентироваться на поиск наименьшей частоты из набора (4).

Для этого введем понятие субполосной корреляции:

$$\varphi^r_{\tau,N} = V^r_{\tau,N} / \sqrt{P_r(\vec{x}_1) P_r(\vec{x}_\tau)} , \tag{8}$$

где [3]

$$V^r_{\tau,N} = \int_{\omega \in \Omega_r} X_1(\omega) X^*_\tau(\omega) d\omega / 2\pi , \tag{9}$$

$$\vec{x}_1 = (x_1, x_2, ..., x_N)^T , \quad \vec{x}_\tau = (x_{1+\tau}, x_{2+\tau}, ..., x_{N+\tau})^T , \tag{10}$$

$$X_1(\omega) = \sum_{i=1}^{N} x_i e^{-j\omega(i-1)} , \quad X_\tau(\omega) = \sum_{i=1}^{N} x_{i+\tau} e^{-j\omega(i-1)} , \tag{11}$$

$$\Omega_r = [-\Omega_{2r}, -\Omega_{1r}) \cup [\Omega_{1r}, \Omega_{2r}) , \tag{12}$$

$$P_r(\vec{y}) = \int_{\omega \in \Omega_r} |Y(\omega)|^2 \, d\omega / 2\pi , \tag{13}$$

$$Y(\omega) = \sum_{i=1}^{N} y_i e^{-j\omega(i-1)}. \quad (14)$$

Подставляя в соотношение (8) и (13) определения (10) и (14), нетрудно получить представления для субполосной корреляции:

$$\varphi^r_{\tau,N} = \vec{x}_1^T A_r \vec{x}_\tau / \sqrt{\vec{x}_1^T A_r \vec{x}_1 \vec{x}_\tau^T A_r \vec{x}_\tau}, \quad (15)$$

где [3]

$$A_r = \{a^r_{ik}\}, \quad i,k=1,\ldots,N$$

$$a^r_{ik} = 2a^0_{ik}\cos(\omega_r(i-k)), \quad (16)$$

$$a^0_{ik} = \sin(\Delta\Omega_r/2(i-k))/(\pi(i-k)), \quad \omega_r = (\Omega_{2r}+\Omega_{1r})/2, \quad (17)$$

$$\Delta\Omega_r = \Omega_{2r} - \Omega_{1r}. \quad (18)$$

Очевидно, что

$$\varphi^r_{\tau,N} \le 1, \quad (19)$$

причем правая часть достигается только при выполнении условия пропорциональности:

$$\vec{x}_\tau = c\vec{x}_1. \quad (20)$$

Легко также понять, что $P_r(\vec{y})$ определяет такое значение энергии отрезка сигнала, попадающей в рассматриваемый частотный интервал. Поэтому наличие в (6) знаменателя позволяет обеспечить чувствительность к свойствам частотных полос с малым уровнем энергии.

Таким образом, определение частоты основного тона сводится к вычислению характеристик (6) при разных значениях τ и ω_r.

При каждом ω_r можно определить такое:

$$M_r = \arg\max \varphi^r_{\tau,N}, \quad 1 \le \tau \le K, \quad (21)$$

которое можно использовать для оценивания относительной частоты основного тона:

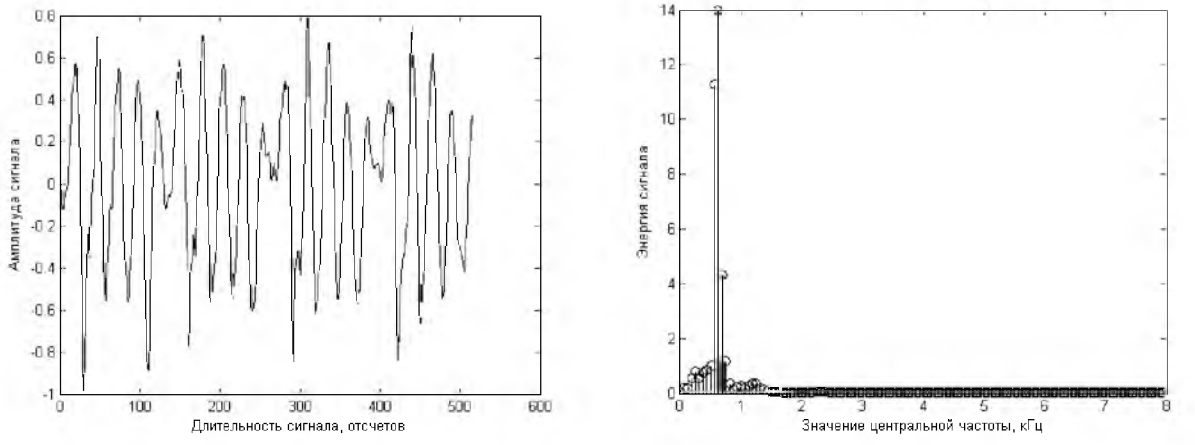
$$\nu_r = 1/M_r, \quad (22)$$

если выполняется условие:

$$\varphi^r_{M_r,N} \ge h \in (0.8 \div 1). \quad (23)$$

Основной интерес представляет оценка при наименьшей возможной частоте ω_r.

Для анализа было использовано следующее разбиение частотной оси на интервалы: первый интервал шириной 62,5Гц имеет начало в точке 0, все последующие имеют ширину 125Гц, причем, центры этих интервалов отстоят друг от друга на $2\pi/N$. На рис. 2 представлено распределение энергии по частотным интервалам при использовании такого разбиения оси частот для фрагмента сигнала, соответствующего звуку «а».

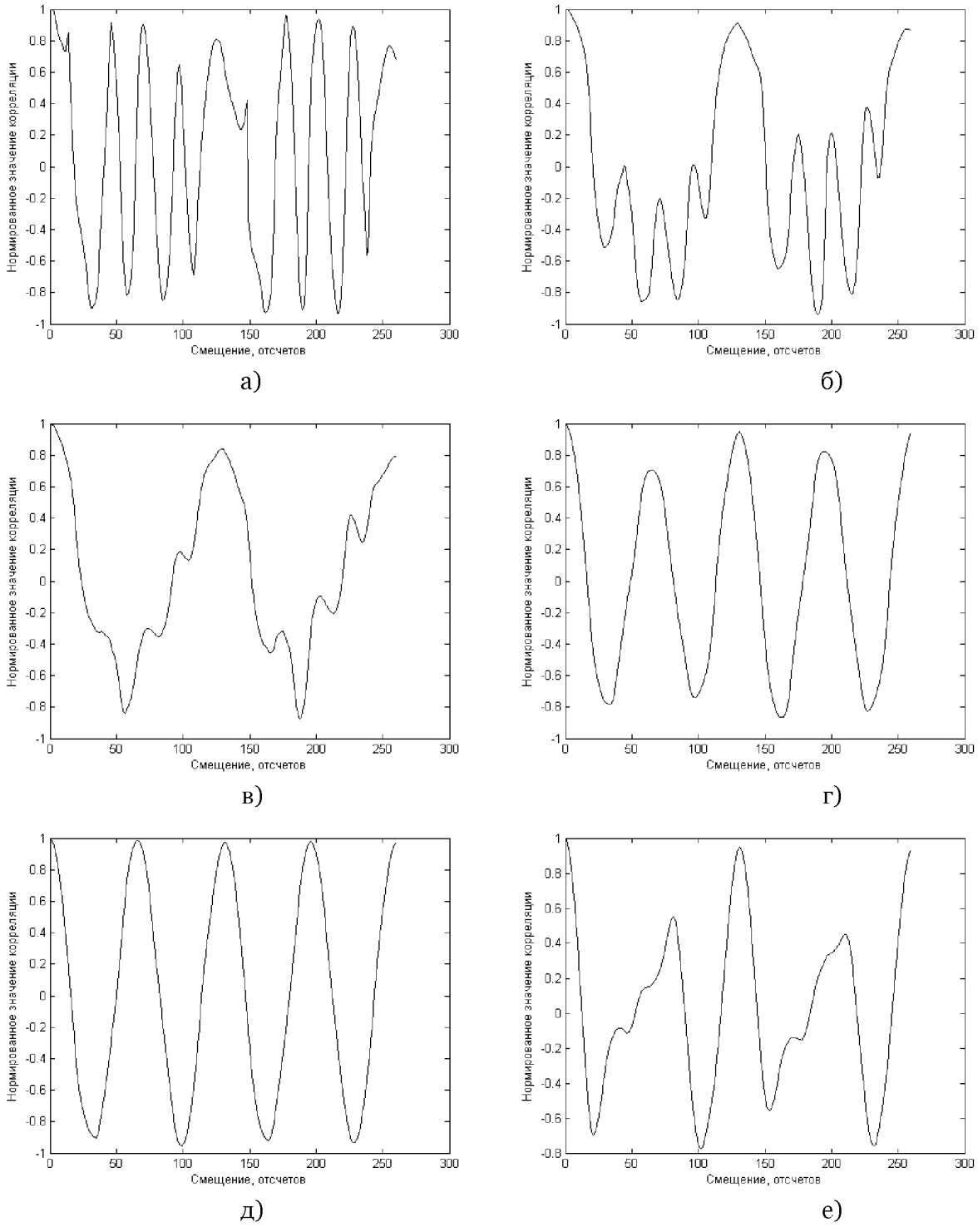


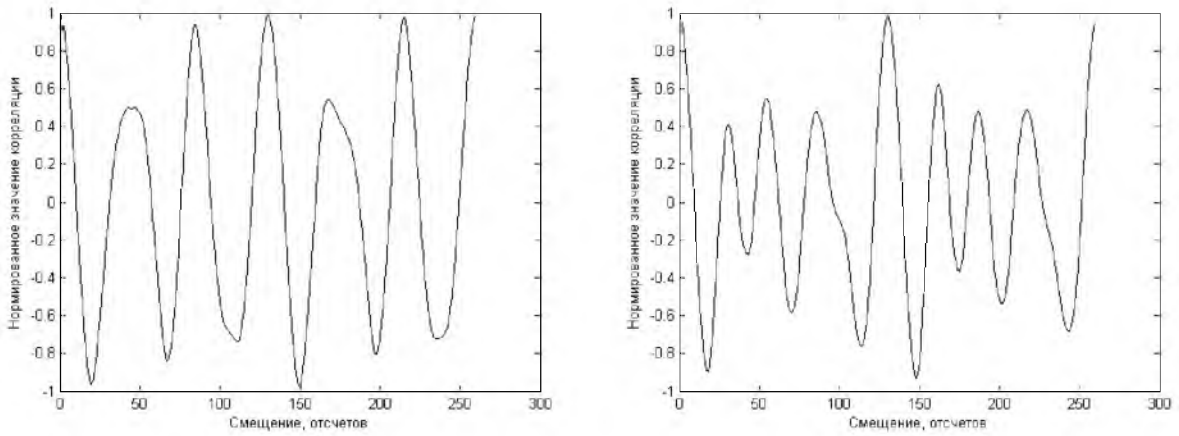
а) б)

Рис. 2. Звук «А»:
а) фрагмент сигнала (f_d=16кГц);
б) распределение энергии по частотным интервалам (N=256 отс, f_d=16кГц)

Для исследования использовался центрированный фрагмент сигнала. Распределение энергии было оценено для отрезка, соответствующего первым 256 отсчетам. Анализ распределения энергии по частотным интервалам, представленный на рисунке 2б, показывает, что основная энергия данного сигнала сосредоточена в диапазоне до 1,5кГц. Наибольшая часть энергии сосредоточена в частотном интервале с центральной частотой равной 625Гц. В свою очередь, частота основного тона анализируемого фрагмента сигнала составляет 124Гц. Таким образом, для данного отрезка сигнала проявляется ситуация, когда максимум правой части выражения (3) наблюдается при значении m в выражении (4) равном 5.

На рис. 3 представлены результаты оценки нормированной корреляции вида (15) для диапазона от 0 до 500 Гц, в который может попасть значение частоты основного тона речевого сигнала.





ж) з)

Рис.3. Нормированные значения корреляции для фрагмента сигнала, соответствующего звуку «А» в частотном интервале: а) (0:62,5)Гц; б) (0:125)Гц; в) (62,5:187,5)Гц; г) (125:250)Гц; д) (187,5:312,5)Гц; е) (250:375)Гц; ж) (312,5:437,5)Гц; з) (375:500)Гц

Анализ результатов экспериментов показывает, что наличие максимума корреляции наблюдается примерно при одном и том же значении смещения окна анализа практически для всех представленных частотных интервалов. Исключение составляют диапазон (0:62,5) Гц и (187,5:312,5) Гц. Причем, поведение функции корреляции в интервале (187,5:312,5) Гц соответствует проявлению частоты 625Гц.

Изменение энергии в частотных интервалах вызвано работой речевого аппарата человека, а также окружающими шумами. Проявление шумов наиболее сильно проявляется в мало энергетических частотных интервалах. Таким образом, для оценивания периода основного тона необходимо учитывать только те частотные интервалы, которые несут основную информацию о речевом аппарате человека. Такие интервалы называются информационными [4, 5]. Для определения информационных частотных интервалов может быть использована частотная концентрация, характеризующая наименьшее количество диапазонов, в которых сосредоточена заданная доля энергии m:

$$f_{NR}^{m} = \min d_{NR}^{m}. \tag{24}$$

Здесь для правых частей выполняется неравенство:

$$\sum_{k=1}^{d_{NR}^{m}} P_{(k)N} \geq m \,\|\, \vec{x}_N \,\|^2 = m \sum_{i=1}^{N} x_i^2, \tag{25}$$

где m – задаваемая доля общей энергии, которая должна быть сосредоточена в указанном минимальном количестве частотных интервалов;

$\vec{x}_N$ – отрезок сигнала, длительностью N отсчетов;

$P_{(k)N}$ – значения энергий в заданных интервалах, после упорядочивания их по убыванию.

Индекс в скобках у слагаемых суммы слева соотношения (25) означает, что части энергий P_{kN} упорядочиваются по убыванию, то есть имеет место

$$P_{(k)N} \in \{P_{rN}, r = 1,..,R\}; P_{(k+1)N} \leq P_{(k)N}, k = 1,..,R. \tag{26}$$

Для принятия решения о значении частоты основного тона предлагается использовать характеристику, представляющую собой зависимость среднего значения коэффициента корреляции среди информационных частотных компонент от соответствующего значения смещения:

$$\varphi_{M_r,N}^{m}(k) = \sum_{r=1}^{f_{NR}^{m}} \varphi_{M_r,N}^{r}(k) / f_{NR}^{m}, \tag{27}$$

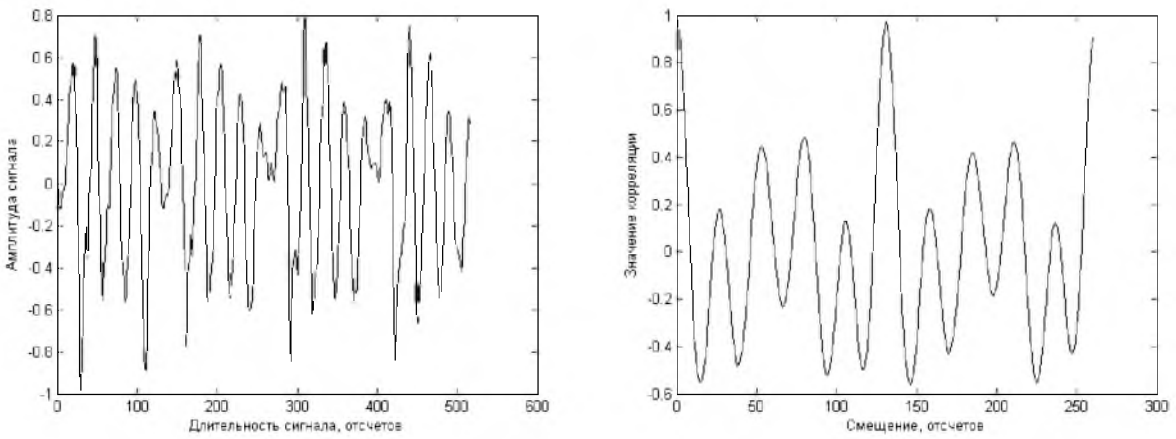
где f_{NR}^{m} – минимальное количество частотных интервалов, в которых сосредоточена заданная доля энергии звукового отрезка;

$\varphi_{M_r,N}^{r}(k)$ – значение корреляции компонент из соответствующих частотных интервалов;

k – смещение окна анализа.

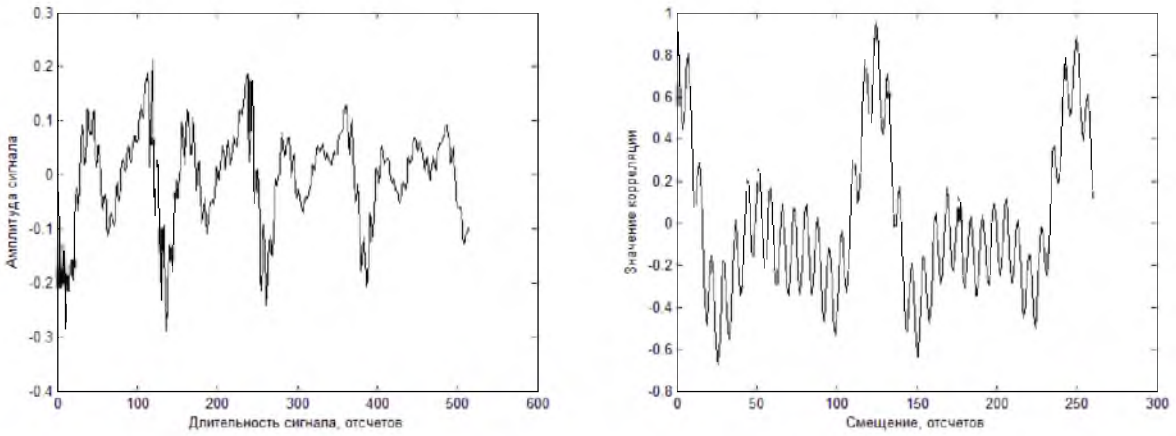
Использование усредненной характеристики позволит учесть поведение сигнала во всех информационных частотных характеристиках. В том случае, если для большинства частотных интервалов максимальное значение корреляции наблюдается для одного и того же значения смещения, то и для усредненной характеристики максимум будет наблюдаться в той же точке.

На рис. 4-7 представлены результаты оценки характеристики вида (27) для фрагментов сигналов, соответствующих некоторым звукам русской речи.



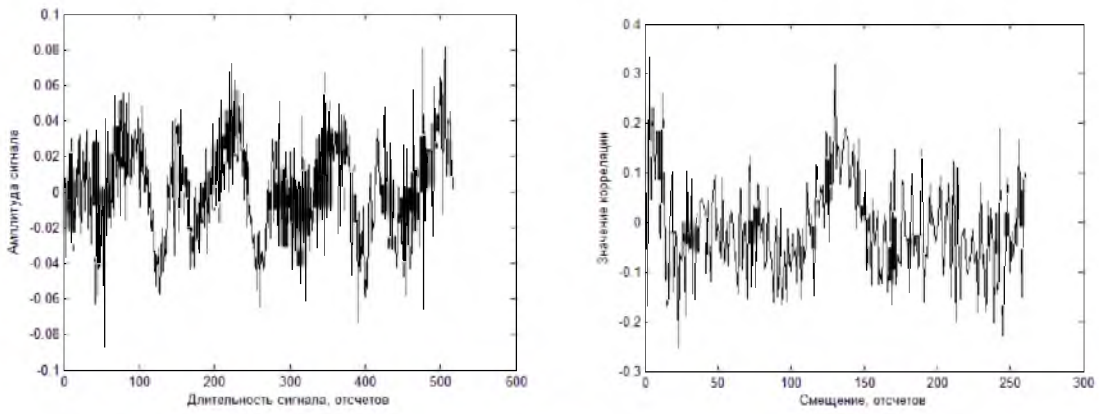
а) б)

Рис. 4. Звук «А»:
а) фрагмент сигнала (N=256, f_d=16кГц);
б) оценка корреляции по информационным частотным интервалам (N=256, m=0,9)



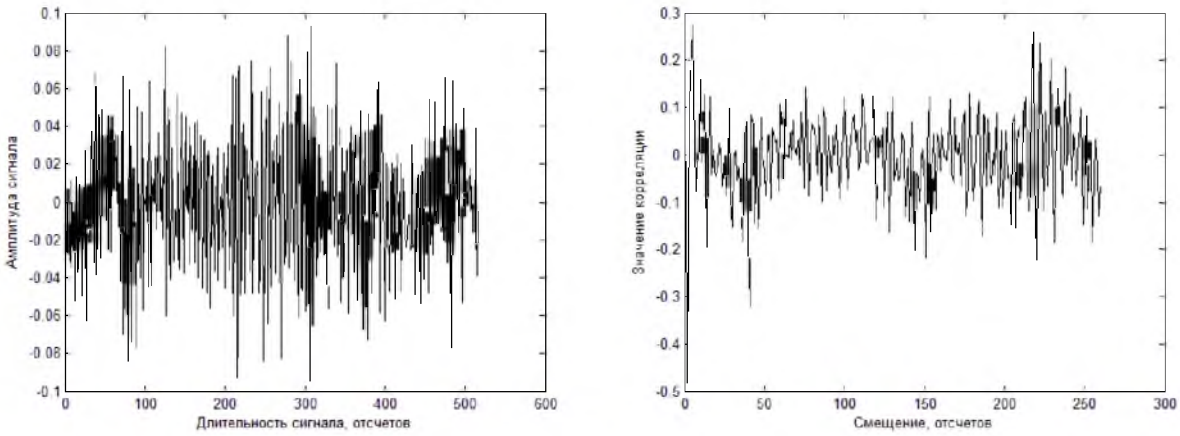
а) б)

Рис. 5. Звук «И»: а) фрагмент сигнала (N=256, f_d=16кГц);
б) оценка корреляции по информационным частотным интервалам (N=256, m=0,9)



а) б)

Рис. 6. Звук «Ж»: а) фрагмент сигнала (N=256, f_d=16кГц);
б) оценка корреляции по информационным частотным интервалам (N=256, m=0,9)



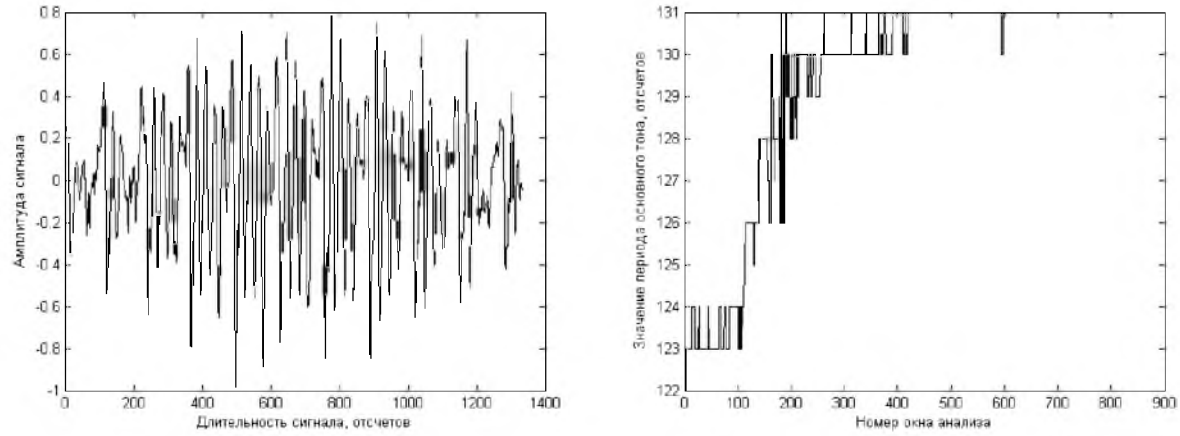
а) б)

Рис. 7. Звук «Ч»: а) фрагмент сигнала (N=256, f_d=16кГц);
б) оценка корреляции по информационным частотным интервалам (N=256, m=0,9)

Анализ результатов показывает, что максимальное значение для звука «А» наблюдается при смещении равном 131 отсчет и достигает 0,97. Эта величина смещения соответствует периоду основного тона анализируемого фрагмента сигнала. Для звука «И» максимальное значение равное 0,96 наблюдается при смещении в 125 отсчетов, что также соответствует периоду основного тона анализируемого отрезка сигнала. Для звуков «Ж» и «Ч» максимальные значения не превышают 0,32, что позволяет определить данные фрагменты как невокализованные участки речевых сигналов. Анализ рисунков также показывает, что для звука «Ж» можно обнаружить наличие ярко выраженного выброса в характеристике при смещении в 130 отсчетов. Одной из особенностей звука «Ж» является участие голосовых связок при его произношении, что проявляется как наличие периодической составляющей на фоне шума.

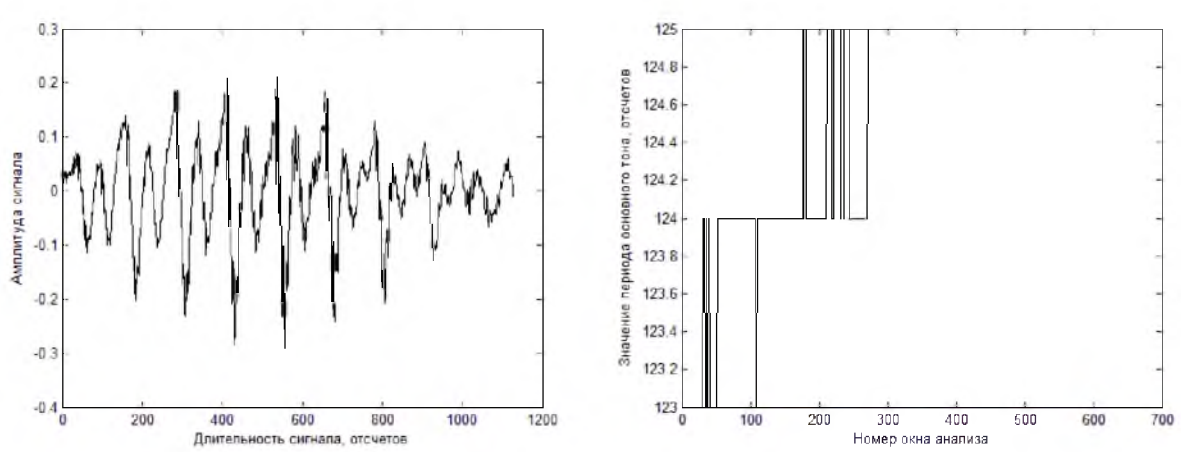
На рис. 8-11 представлены фрагменты сигналов, соответствующих некоторым звукам русской речи, и результаты оценки периода основного тона на основе правила:

$$M = \arg\max \varphi^{m}_{M_p,N}(k), \quad 1 \le \tau \le K . \tag{28}$$



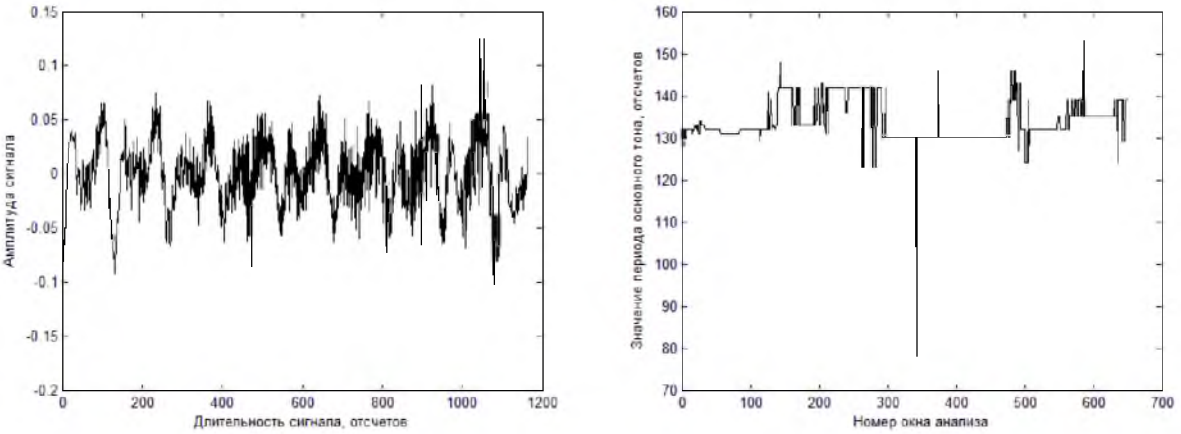
а) б)

Рис. 8. Звук «А»: а) фрагмент сигнала (N=256, f_d=16кГц);
б) результат оценки периода основного тона (N=256, m=0,9)



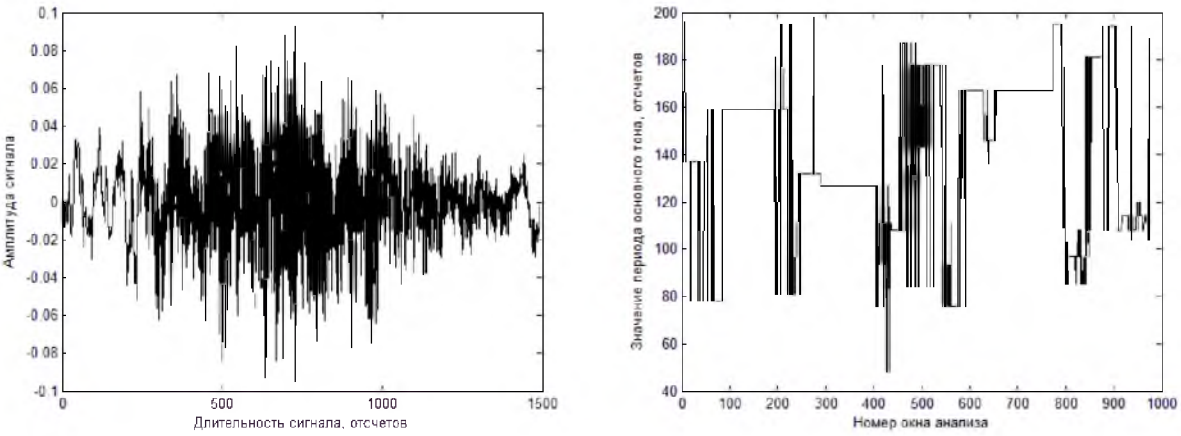
а) б)

Рис. 9. Звук «И»: а) фрагмент сигнала (N=256, f_d=16кГц);
б) результат оценки периода основного тона (N=256, m=0,9)



а) б)

Рис. 10. Звук «Ж»: а) фрагмент сигнала (N=256, f_d=16кГц);
б) результат оценки периода основного тона (N=256, m=0,9)



а) б)

Рис. 11. Звук «Ч»: а) фрагмент сигнала (N=256, f_d=16кГц);
б) результат оценки периода основного тона (N=256, m=0,9)

Анализ представленных результатов показывает, что для гласных звуков разброс принятых значений периода основного тона не превышает 7%. Для звука «Ж» без учета значения 78 отсчетов на 342 окне анализа разброс составляет порядка 20%. Для звука «Ч» разброс значений существенно больше, что свойственно шумным звукам речи.

Таким образом, использование метода, основанного на учете корреляции компонент, соответствующих информационным частотным интервалам, позволяет достаточно

точно определять период основного тона речевых сигналов. Данный метод может быть также использован для определения вокализованных и невокализованных фрагментов речевого сигнала на основе анализа максимальных значений коэффициентов корреляции и стабильности поведения значений периода основного тона.

Исследования выполнены при поддержке проекта № 8.2251.2011.

Список литературы

1. Рабинер, Л.Р. Цифровая обработка речевых сигналов /Л.Р. Рабинер, Р.В. Шафер – М.: Радио и связь, 1981. – 496с.
2. Михайлов В.Г., Златоустова Л.В. Измерение параметров речи /под.ред. М.А.Сапожкова – М.: Радио и связь, 1987. – с.168
3. Жиляков Е.Г. Вариационные методы анализа сигналов на основе частотных представлений / Е.Г. Жиляков, С.П. Белов, А.А. Черноморец // Журнал «Вопросы радиоэлектроники», серия ЭВТ, 2010 г, с.10-26.
4. Фирсова А.А. О различии концентрации энергии по частотным диапазонам на отрезках сигналов, соответствующих шипящим звукам русской речи и шумам / А.С. Белов, А.В. Курлов, А.А. Фирсова // журнал «Научные ведомости БелГУ» Серия: История. Политология. Экономика. Информатика, №13(108) 2011, выпуск 19/1, 2011 г, с186-190.
5. Белов С.П. О различиях частотных свойств информационных и неинформационных звуковых сигналов речевого диапазона / С.П. Белов, А.С. Белов // журнал «Научные ведомости БелГУ» Серия: История. Политология. Экономика. Информатика, №10(50) 2008, выпуск 8/1, 2008 г, с86-93.

ESTIMATION THE PITCH PERIOD SOUNDS RUSSIAN SPEAKING

E.G. ZHILYAKOV
A.A. FIRSOVA

Belgorod National Research University

e-mail: zhilyakov@bsu.edu.ru

The paper presents a new method of estimation pitch period based on the use analysis and evaluation of the correlation component of each frequency band.

Keywords: speech signal, analysis of the speech signal, the frequency of the fundamental, analysis.