ОБ ОБРАБОТКЕ РЕЧЕВЫХ СИГНАЛОВ

Жиляков Е.Г., Байдиков А.Н.

Университет потребительской кооперации
308023, г. Белгород ул. Садовая 116а, каф. информационных систем и технологий
(0222) 26-38-31, Fax (0222) 26-49-65, e-mail bupk@intbel.ru

Abstract. Предложен метод сегментации и оценки периода основного тона речевых сигналов. Произведен сравнительный анализ результатов метода с ранее известными методами оценки периода основного тона.

1. Представление речевых сигналов

В работе [1] было преложено новое представление речевых сигналов. Это представление имеет следующий вид

$$f(t) = \sum_{k=-\infty}^{\infty} Z_k(T,t) \cdot S(\frac{t}{2T} - k), \tag{1}$$

где

$$Z_k(T,t) = \sum_{n=-\infty}^{\infty} f(t-2nT) \cdot S(\frac{t}{2T} - n - k), \tag{2}$$

периодические функции с периодом $2T$, то есть имеет место

$$Z_k(T, t + 2mT) = Z_k(T, t), \quad m = \pm 0, \pm 1, \pm 2 \pm ..., \tag{3}$$

k – номер функции, и весовое окно

$$S(x) = \frac{\sin \pi x}{\pi x}. \tag{4}$$

Доказательство справедливости этого представления приведено в [5], для непрерывных, ограниченных вещественных функций непрерывного аргумента t, удовлетворяющих условиям:

$$\int_{-\infty}^{\infty} |f(t)| dt < \infty, \tag{5}$$

$$\frac{|f(t)|}{|t^{\varepsilon}|} \le L < \infty, \tag{6}$$

где ε – произвольное малое положительное число $\varepsilon > 0$.

Периодические функции $Z_k(T, t)$, называются *структурными* функциями [1]. Так же в [1] было указано соотношение (2), которое определяет некоторую процедуру весовой обработки. При этом свойства весовой функции (4) таковы, что будут иметь место равенства:

$$Z_k(T,t) = f(2kT) \tag{7}$$

а в остальных точках периода приближенно выполняется

$$|Z_k(T,t) - f(t+2kT)| \approx 0, \quad 0 \le t \le 2T. \tag{8}$$

Иными словами структурные функции в основном содержат информацию о локальных свойствах сигнала в соответствующих интервалах (смежных с периодом $2T$) числовой оси [1].

Для процесса речеобразования резонансные частоты трубы голосового тракта принято называть формантными частотами или просто *формантами* [3]. Формантные частоты зависят от конфигурации и размеров голосового тракта на определенный момент времени.

Процесс речеобразования принято представлять [2, 3] как отклик фильтра голосового тракта на возбуждающие импульсы голосовых связок. Для вокализованных звуков возбуждающая последовательность является квазипериодической. Поэтому для таких звуков существует еще оценка называемая *периодом основного тона* или *частотой* основного тона.

В формировании выходного речевого сигнала участвуют затухающие квазипериодические колебания, частота которых, соответствует формантным частотами голосового тракта или формантам [2, 3]. В силу этого следствия представление (1) в некоторой степени сопоставимо с моделью речеобразования описанной в [2, 3], в виду наличия весового окна (4), которое соответствует природе (затухающее и периодическое) процесса речеобразования. Путем подбора периода структурных функций $2T$ можно произвести оценку формантных частот голосового тракта и период основного тона речевого сигнала.

2. Сегментация и оценивание периода основного тона речевых сигналов

Под сегментацией речевых сигналов понимается процесс разбиения всего сигнала на интервалы, обладающие некоторыми одинаковыми свойствами, примерно равной мощностью энергии или наличием квазипериодической составляющей проявляющейся с некоторым периодом во всем сигнале. Процедура сегментации состоит из двух этапов: *на первом* этапе строится функционал; *на втором* проводится анализ полученной функциональной зависимости.

Первый этап – построение функционала.

Выбирается диапазон изменения периода структурных функций $T=[T_{min}, T_{max}]$; для каждого значения T_i, i = (1.. Θ) производится вычисление значения функционала, Θ – количество измерений функционала (мощность результирующего вектора значений функционала).

Функционал выбирается из предпосылки выполнения условия (8). Иными словами качество выполнения условия (8) для определенного периода структурных функций $2T$, характеризуется близостью структурных функций к исходному речевому сигналу на интервале $[2kT, 2(k + 1)T]$

$$Q(T) = \sqrt{\frac{D_q(T) + M_q^2(T)}{F_t(T)}}. \qquad (9)$$

Функционал (9) описывает зависимость между периодом структурных функций (2) и среднеквадратическим отклонением структурных функций от исходного речевого сигнала.
где

$$F_t(T) = \frac{1}{R}\sum_{k=0}^{R-1}\sum_{t=0}^{T-1} f^2(t + 2kT), \qquad (10)$$

$F_t(T)$ – среднее значение энергии сигнала, не зависит от T, так как $R = L / T$,
L –общая длина речевого сигнала.

Ввиду конечности сигнала [6] и наличия весового окна (4) в (1) и (2), необходимо выполнить переход к ограниченным пределам суммирования рядов (1) и (2). Общее количество структурных функций $M = R + O$, где
R – кол-во периодов структурных функций укладывающихся на всей длине сигнала L;
O – кол-во периодов структурных функций перекрывающее границы сигнала.

Наличие параметра O вызвано использованием весового окна (4) в (2), что вызывает эффект захвата энергии сигнала из соседних интервалов. Следовательно, отбрасывая такие структурные функции, часть энергии сигнала на концах сигнала будет потеряна, что является нежелательным фактором для точности восстановления сигнала по структурным функциям.

На (рис 2.1) представлены совмещенные во времени периоды трех разных структурных функций полученные по (2) для речевого сигнала (буква «а»). Структурные функции практически в точности совпадают с исходным сигналом. Структурные функции на рисунке имеют заостренные пики.

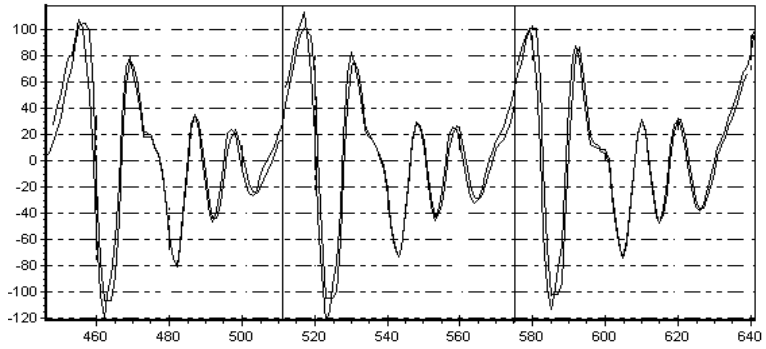


Рис 2.1 *Структурные функции*, наложенные на исходный речевой сигнал. Минимум функционала $Q_{\min} = 0{,}23$; при периоде $T = 64$ *smpl* – отсчета, частота дискретизации сигнала 11кГц.

Введем оценочный вектор $q(T)$, необходимый для выполнения условия (8)

$$q_k(T) = \sqrt{\sum_{t=0}^{T-1}(Z_k(T,t) - f(t + 2kT))^2}, \qquad (11)$$

$q_k(T)$ – элемент оценочного вектора $q_k(T)$, среднеквадратическое отклонение k-ой структурной функции от исходного сигнала на интервале $[2kT, 2(k + 1)T]$;

$$Z_k(T, t) = \sum_{n=N\min}^{N\max} f(t - 2nT)\cdot S(\frac{t}{2T} - n - k), \qquad (12)$$

где $N_{min} = M_{min}\ 2T$, $N_{max} = M_{max}\ 2T$ – границы речевого сигнала; M_{min} *и* M_{max} – минимальный и максимальный номер структурной функции соответственно.

$$M_q(T) = \frac{1}{R}\sum_{k=0}^{R-1} q_k(T), \qquad (13)$$

$M_q(T)$ – математическое ожидание оценочного вектора $q_k(T)$;

$$D_q(T) = \frac{1}{R}\sum_{k=0}^{R-1}(q_i(T) - M_q(T))^2 \qquad (14)$$

$D_q(T)$ – дисперсия оценочного вектора $q_k(T)$;

Второй этап – исследование функционала:

Очевидно, что минимумы функционала являются минимальным среднеквадратическим отклонением структурных функций от речевого сигнала. Следовательно, при таких значениях периода $2T$ выполняется условие (8), структурные функции (12) наиболее близки в среднеквадратическом смысле к исходному сигналу. Иначе говоря, наблюдается выделение некоторой периодической составляющей речевого сигнала.

Рассмотрим несколько примеров вычисления функционала (9) для различных речевых сигналов. На (рис 2.2 – 2.5) представлены функционалы, вычисленные для коротких речевых сигналов, которые представляют собой отдельно произнесенные буквы алфавита (частота дискретизации сигналов 11кГц).

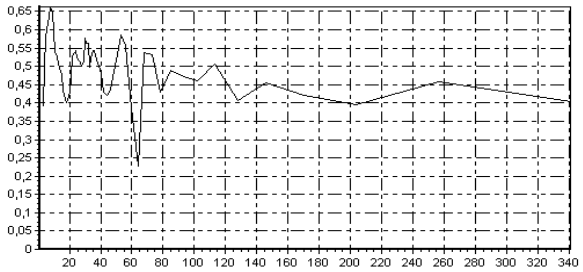


Рис 2.2 Буква «а»

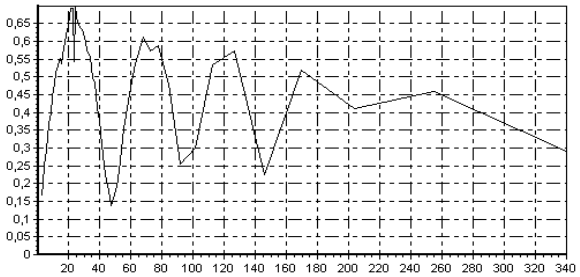


Рис 2.3 Буква «и»

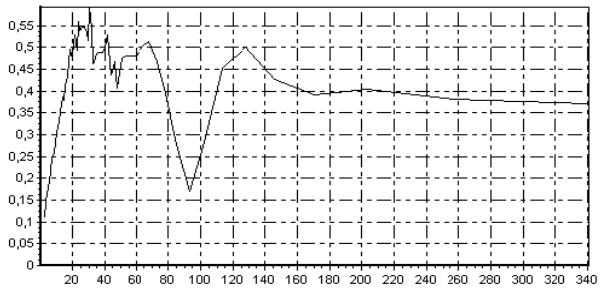


Рис 2.4 Буква «м»

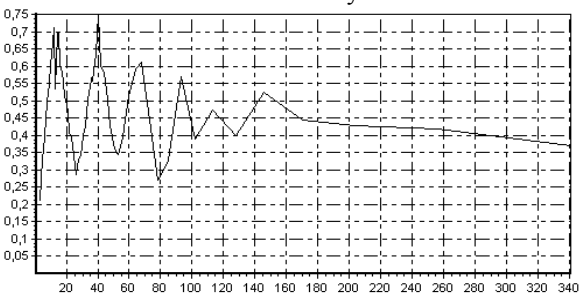


Рис 2.5 Буква «о»

Следует отметить, что при анализе полученных функционалов область малой длительности периода структурных функций (до 5 *smpl* –отсчетов (11кГц)), рассматривать не следует, так как
при $T \to 0 \Rightarrow Q \to 0$.
Анализируя (рис 2.2 – 2.5) получим минимумы функционалов (табл. 2.1).

Речевой сигнал	Q_{min}	***T**, smpl*	*≈**T**, мс*	*≈**F**, Гц*
	1	2	4	3
Буква «а»	0,23	64	6	172
Буква «и»	0,14	48	4	230
Буква «м»	0,16	93	8	119
Буква «о»	0,27	78	7	141

Табл. 2.1 Минимумы функционалов

Q_{min} – мин. значение функционала;
T – значение периода в (*smpl* и *мс*);
$F = 1 / T$ – частота периода в (*Гц*);

Из полученных результатов и анализа [2, 3] можно сделать вывод о том, что полученные периоды структурных функций являются периодом или частотой основного тона представленных речевых сигналов. Этот факт также можно наблюдать на (рис 2.1).

Другие минимумы функционала характеризуют периоды кратные периоду основного тона сигнала либо другие периодические составляющие сигнала, например форманты.

Так же были проведены эксперименты по сегментации с отдельно произнесенными словами, фразами и предложениями.

Значение функционала для слова “Вадим” имеет минимум Q_{min} = 0,4 при T = 67 *smpl*, что соответствует усредненному периоду основного тона для всего слова и Q_{min} = 0,4 при T = 1281 *smpl*, что соответствует средней длине буквы в этом слове.

На фразе “Вадим шел в дом” функционал (9) имеет минимальное значение Q_{min} = 0,2 при T = 6406 *smpl*, что соответствует средней длине слова из этой фразы.

Исходя из этих результатов, можно сделать вывод о том, что эта процедура сегментации выделяет периоды слитно - произнесенных участков речевого сигнала и его периодические составляющие. Для фраз или предложений аргумент минимума функционала определяет среднюю длительность каждого слова и соответственно их количество. Для слов, длительность букв его составляющих, их количество. Для букв, период основного тона, другие периодические составляющие.

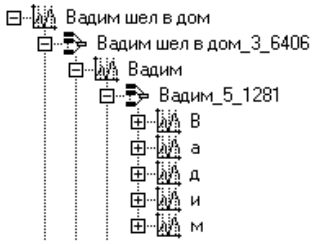


Рис 2.6 Иерархическая сегментация

Обобщая вышеприведенные заключения можно предложить *иерархический метод сегментации* речевых сигналов. Суть, которого состоит в последовательном многоуровневом анализе речевого сигнала (рис.2.6). Вначале находится глобальный минимум функционала (9) для всего сигнала. Далее весь сигнал разбивается на интервалы в соответствии с выделенным периодом. После этого вся процедура повторяется для каждого интервала как для всего сигнала. В результате мы получаем иерархически разбитые интервалы сигнала, которые соответствуют составляющим элементам сигнала, описание которым было дано выше.

3. Сравнительный анализ метода с другими известными методами

Для сравнительного анализа используемого метода сегментации использовались методы определения периода основного тона сигнала: автокорреляционный метод и метод с использованием кепстрального анализа. Результаты сравнения методов по оценке периода основного тона приведены в (табл. 3.1)

Речевой сигнал	**Автокорреляционный метод**		**Кепстральный анализ**		**Метод структурных функций**	
	***T**, мс*	***F**, Гц*	***T**, мс*	***F**, Гц*	***T**, мс*	***F**, Гц*
Буква «а»	6	178	6	178	6	172
Буква «и»	5	221	4	230	4	230
Буква «м»	9	114	8	113	8	119
Буква «о»	7	141	7	140	7	141

Табл. 3.1 Сравнительный анализ методов оценки периода основного тона

По результатам работы методов (табл. 3.1) можно сделать вывод о том, что метод структурных функций с использованием функционала вида (9) дает адекватную оценку периода основного тона. Погрешность оценки в сравнении с вышеприведенными методами составляет ≈1мс (период) или ≈11Гц (частота) основного тона речевого сигнала.

Библиография

1. Жиляков Е.Г., Байдиков А.Н., Описание речевых сигналов на основе параметрического представления / Материалы II – ой международной конференции-выставки “Цифровая обработка сигналов и ее применения – DSPA’99”. Москва, 1999г., с553-557
2. Рабинер Л.Р., Шафер Р. В. Цифровая обработка речевых сигналов. М.: Радио и связь, 1981
3. Маркел Д. Д., Грей А.Х. Линейное предсказание речи. М.: Связь, 1980
4. Обнаружение изменения свойств сигналов и физических систем / Под ред. М. Басевиль, А. Банвениста. – М.: Мир, 1989
5. Жиляков Е.Г., Тубольцев М. Ф. Об одном новом представлении функций и его применениях в задачах обработки сигналов / Материалы научно- технической конференции “Направления развития систем и средств радиосвязи”. Воронеж, 1996
6. Хургин Я. И., Яковлев В. П., Финитные функции в физике и технике. М.: Наука, 1971

ABOUT PROCESSING OF THE SPEECH SIGNALS

Zhyljakov E.G., Baidikov A.N.

University of the consumer cooperatives

308023, Belgorog Sadovaja st.116a, chair of the information systems and technology

(0222) 26-38-31, Fax (0222) 26-49-65, e-mail bupk@intbel.ru

Abstract For the speech signals offer segmentation method and period of pitch estimation. Derive comparative analysis of the method results with the above known period pitch estimation methods.

1. Speech signals presentation

In the work [1] was offer new presentation of the speech signals. This presentation has next view

$$f(t) = \sum_{k=-\infty}^{\infty} Z_k(T,t) \cdot S\left(\frac{t}{2T} - k\right), \tag{1}$$

where

$$Z_k(T,t) = \sum_{n=-\infty}^{\infty} f(t-2nT) \cdot S\left(\frac{t}{2T} - n - k\right), \tag{2}$$

Periodical functions with the period $2T$, therefore take place

$$Z_k(T, t + 2mT) = Z_k(T, t), \quad m = \pm 0, \pm 1, \pm 2 \pm ..., \tag{3}$$

k – function number, and the weight function

$$S(x) = \frac{\sin \pi x}{\pi x}. \tag{4}$$

The proof of justice this presentation adduced in [1], for continuous, limited real functions of the continuous argument t, answer this conditions

$$\int_{-\infty}^{\infty} |f(t)| dt < \infty, \tag{5}$$

$$\frac{|f(t)|}{|t^{\varepsilon}|} \le L < \infty, \tag{6}$$

where ε – arbitrary small positive number $\varepsilon > 0$.

Periodical functions $Z_k(T, t)$, called by *structure* functions [1]. Also in [1] was designate relation (2), witch determine some procedure of the weight processing. Herewith properties of the weight function (4) such, what will take place equations

$$Z_k(T,t) = f(2kT) \tag{7}$$

and in other points of period approximately performs

$$|Z_k(T,t) - f(t+2kT)| \approx 0, \quad 0 \le t \le 2T. \tag{8}$$

Reword, structure functions in the main contain information about the local properties of the signal in coincide intervals (neighbouring with period $2T$) of the number axis [1].

In formation out speech signal be in damping quasi-periodical oscillation, witch frequency, check with formant frequency of the vocal tract either formants [2, 3]. In mean of this consequence presentation (1) in some degree comparable with model of the speech formation described in [2, 3], in aspect of presence weight function (4), witch coincide kind (damping and periodical) of speech formation process. By way of selection period of the structure functions $2T$ possible perform estimation of the formant frequencies of the vocal tract and pitch of the speech signal period.

2. Segmentation and period pitch estimation of the speech signals

Segmentation procedure consists from two stages: *on first* stage build functional; *on second* perform analysis given functional relation.

First stage – building functional.

Select the range of the period measurement structure functions $T=[T_{min}, T_{max}]$; for each value T_i, $i = (1.. \Theta)$ perform calculation functional value, Θ – quantity of functional measurement

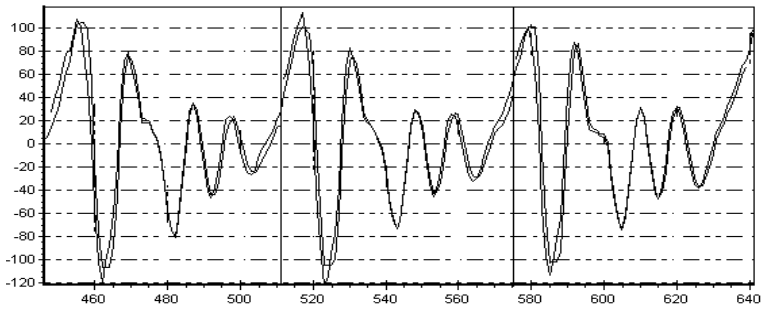
$$Q(T) = \sqrt{\frac{D_q(T) + M_q^2(T)}{F_t(T)}}. \tag{9}$$

Functional (9) where

$$F_t(T) = \frac{1}{R} \sum_{k=0}^{R-1} \sum_{t=0}^{T-1} f^2(t + 2kT), \tag{10}$$

$F_t(T)$ – mean value of signal energy is undependable from T, because $R = L / T$,
L –whole length of the speech signal.

Necessary perform passing to limited ranges in row (1) and (2) recapitulation, because the signal is limited and has weight function (4) in (1) and (2). Whole quantity of structure functions $M = R + O$, where
R – quantity of the structure function period pack on all length of the signal L;
O – quantity of the structure function period overlap ranges of the signal.



Pic 2.1 *Structure function*, witch claped on original signal. Functional minimum $Q_{\min} = 0{,}23$; for period $T = 64$ *smpl*, signal sampling frequency 11кГц.

Bring in mark vector $q(T)$, necessary for performing condition (8)

$$q_k(T) = \sqrt{\sum_{t=0}^{T-1} (Z_k(T,t) - f(t + 2kT))^2}, \tag{11}$$

$$M_q(T) = \frac{1}{R}\sum_{k=0}^{R-1} q_k(T),\ D_q(T) = \frac{1}{R}\sum_{k=0}^{R-1} (q_i(T) - M_q(T))^2. \tag{12}$$

Second stage – functional exploring for local minimums. (Example of the speech signal analysis presents below)

Speech signal (russian)	Q_{min}	***T**, smpl*	*≈**T**, ms*	*≈**F**, Hz*
	1	2	4	3
Letter «а»	0,23	64	6	172
Letter «и»	0,14	48	4	230
Letter «м»	0,16	93	8	119
Letter «о»	0,27	78	7	141

Table 2.1 Functional minimums

Q_{min} – functional min. value;
T – period value (*smpl* and *ms*);
$F = 1 / T$ – period frequency (*Hz*);

Method of the structure function was compared with autocorrelation method and cepstral analysis

Speech signal (russian)	**Autocorrelation method**		**Cepstral analysis**		**Method of the structure function**	
	***T**, ms*	***F**, Hz*	***T**, ms*	***F**, Hz*	***T**, ms*	***F**, Hz*
Letter «а»	6	178	6	178	6	172
Letter «и»	5	221	4	230	4	230
Letter «м»	9	114	8	113	8	119
Letter «о»	7	141	7	140	7	141

Table 3.1 Comparative analysis of the pitch period estimation methods

By results of the methods (table 3.1) possible make conclusion about the structure function method with functional view (9) using give right pitch period estimation. Method fallibility ≈1ms (period) or ≈11Hz (frequency) of the speech signal pitch period.

Bibliography

1. Zhyljakov E.G., Baidikov A.N. Speech signals description based on the parametrical presentation. / The 2nd International conference DIGITAL SIGNALS PROCESSING AND ITS APPLICATIONS, Moscow 1999
2. Rabiner L.R., Schafer R.W. Digital processing of speech signals. Prentice-Hall, Inc., Englewood Cliffs, New Jersey 07632
3. Markel J. D., Gray A.H. Linear prediction of speech. Springer-Verlag Berlin Heidelberg New York 1976