

# Метод оценки частоты основного тона в условиях помех<sup>1</sup>

Гай В. Е.

## Введение

Оценка частоты голоса – одна из основных задач цифровой обработки речевых сигналов. Результаты её решения используются при распознавании и сжатии речи, идентификации / верификации диктора и т.д. Алгоритмы вычисления частоты основного тона нашли своё применение при настройке музыкальных инструментов, создании нотной записи и классификации музыкальных произведений.

Методы, предназначенные для решения данной задачи, анализируют звуковой сигнал в частотной и временной областях. Во временной области оценка частоты осуществляется по количеству пересечений сигналом нулевого уровня [2], по автокорреляционной функции [3, 4], по функции среднего значения разности амплитуд [5]. Алгоритмы, обрабатывающие данные в частотной области, анализируют гармоники спектра [6], используют кепстральный анализ [7]. В [8] рассматривается подход к оценке частоты основного тона, выполняющий одновременную обработку сигнала в частотной и временной областях.

Предлагаемый метод, в качестве базиса, использует теорию активного восприятия (ТАВ), которая была предложена профессором В. А. Утробиним [9]. В рамках ТАВ звуковой сигнал рассматривается как системное образование. Для обнаружения системных элементов используется интегральное преобразование, а для выявления связей между элементами – пространственное дифференцирование. Результатом выявления дифференциальной структуры является спектральное описание сигнала.

Теория активного восприятия не ограничивается только формированием спектрального представления сигнала. В ТАВ входит раздел «Алгебра групп», посвящённый анализу зависимостей между спектральными коэффициентами разложения. Обнаруженные зависимости допускают своё использование на этапах принятия решения и понимания анализируемого звукового сигнала [9]. Базовым преобразованием ТАВ является  $U$ -преобразование.

## 1. $U$ -преобразование

Пусть  $f(t)$  – анализируемый звуковой сигнал, наблюдаемый на конечном отрезке времени. Результат применения  $U$ -преобразования к сигналу  $f$  – многоуровневое (грубо-точное) спектральное представление  $D = \{d_{ij}\}$ ,  $i = \overline{1, K}$ ,  $j \in \overline{1, M_i}$ , где  $K$  – число уровней разложения,  $M_i$  – количество сегментов сигнала на  $i$ -ом уровне разложения,  $d_{ij}$  – спектр, включающий  $L$  спектральных коэффициентов (число используемых фильтров),  $d_{ij}\{k\}$  –  $k$ -ый спектральный коэффициент ( $k = \overline{1, L}$ ),  $f_{ij}$  – сегмент сигнала  $f$ , по которому вычислен спектр  $d_{ij}$ .

Рассмотрим процесс вычисления двухуровневого  $U$ -представления тестового сигнала (сигнал включает 16 отсчётов).

Уровень 1. Интегрирование. Сигнал разбивается на четыре части, амплитуды отсчётов, отнесённые к каждой части складываются (см. рис. 1, части отделены пунктиром). В результате, генерируется массив сумм:  $S = \{12, 13, 8, 10\}$ .

---

<sup>1</sup> Работа выполнена при поддержке Министерства образования и науки РФ в рамках договора № 02.G25.31.0061 от «12» февраля 2013 года (в соответствии с Постановлением Правительства Российской Федерации от 9 апреля 2010 г. № 218).

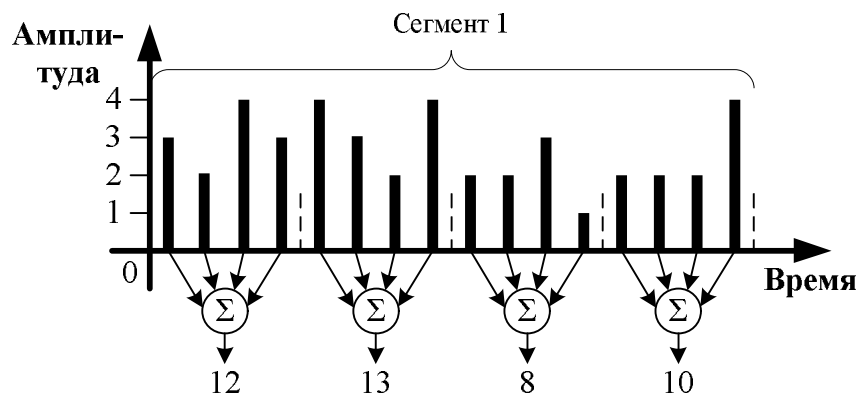


Рис. 1. Интегрирование сигнала (уровень 1)

Уровень 1. Дифференцирование. Применяя к массиву сумм  $S$  фильтры, показанные на рис. 2, вычисляются значения спектральных коэффициентов. Для этого каждый фильтр поэлементно умножается на массив сумм. Например, для второго фильтра в результате умножения будет сгенерирован следующий массив:  $S = \{-12, -13, 8, 10\}$ . Далее выполняют сложение элементов массива:  $(-12) + (-13) + 8 + 10 = -7$ . Результат сложения представляет собой спектральный коэффициент (см. табл. 1).

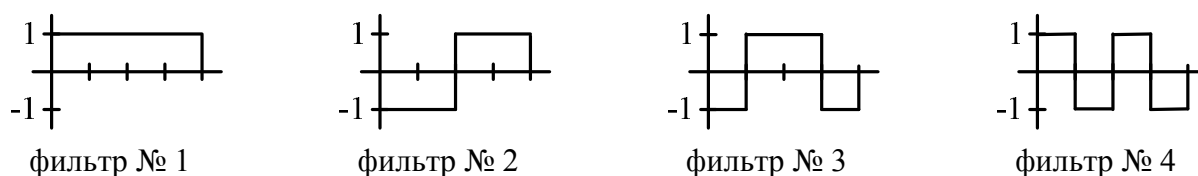


Рис. 2. Базисные функции

Таблица 1. Вычисление спектральных коэффициентов первого уровня

Результат умножения				Фильтр	Спектральные коэффициенты
12	13	8	10	№ 1	43
-12	-13	8	10	№ 2	-7
-12	13	8	-10	№ 3	-1
12	-13	8	-10	№ 4	-3

Уровень 2. Интегрирование. Сигнал разбивается на четыре сегмента, каждый сегмент разбивается на 4 части, вычисляется сумма отсчётов, которые относятся к каждой части (см. рис. 3).

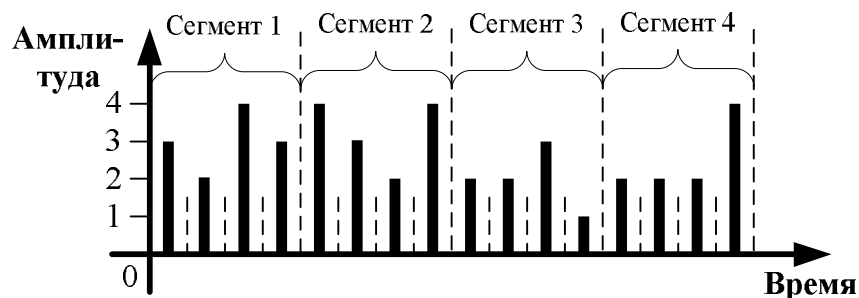


Рис. 3. Интегрирование сигнала (уровень 2)

Уровень 2. Дифференцирование. С помощью фильтров, приведённых на рис. 2 и результатов интегрирования, вычисляются значения спектральных коэффициентов (см. табл. 2).

Таблица 2. Вычисление спектральных коэффициентов второго уровня

Результат умножения								Фильтр	Спектральные коэффициенты	
сегмент 1				сегмент 2					сегмент 1	сегмент 2
3	2	4	3	4	3	2	4	№ 1	12	13
-3	-2	4	3	-4	-3	2	4	№ 2	2	-1
-3	2	4	-3	-4	3	2	-4	№ 3	0	-3
3	-2	4	-3	4	-3	2	-4	№ 4	2	-1
сегмент 3				сегмент 4					сегмент 3	сегмент 4
2	2	3	1	2	2	2	4	№ 1	8	10
-2	-2	3	1	-2	-2	2	4	№ 2	0	2
-2	2	3	-1	-2	2	2	-4	№ 3	2	-2
2	-2	3	-1	2	-2	2	-4	№ 4	2	-2

Описанный способ вычисления  $U$ -преобразования может быть применён к звуковому сигналу произвольной длины.

В приведенном примере для вычисления спектрального представления сигнала используется четыре фильтра. Однако для решения задачи оценки частоты основного тона требуется более тонкий анализ сигнала. В связи с этим, при построении спектрального представления сигнала будет использовано 16 фильтров (см. рис. 4).

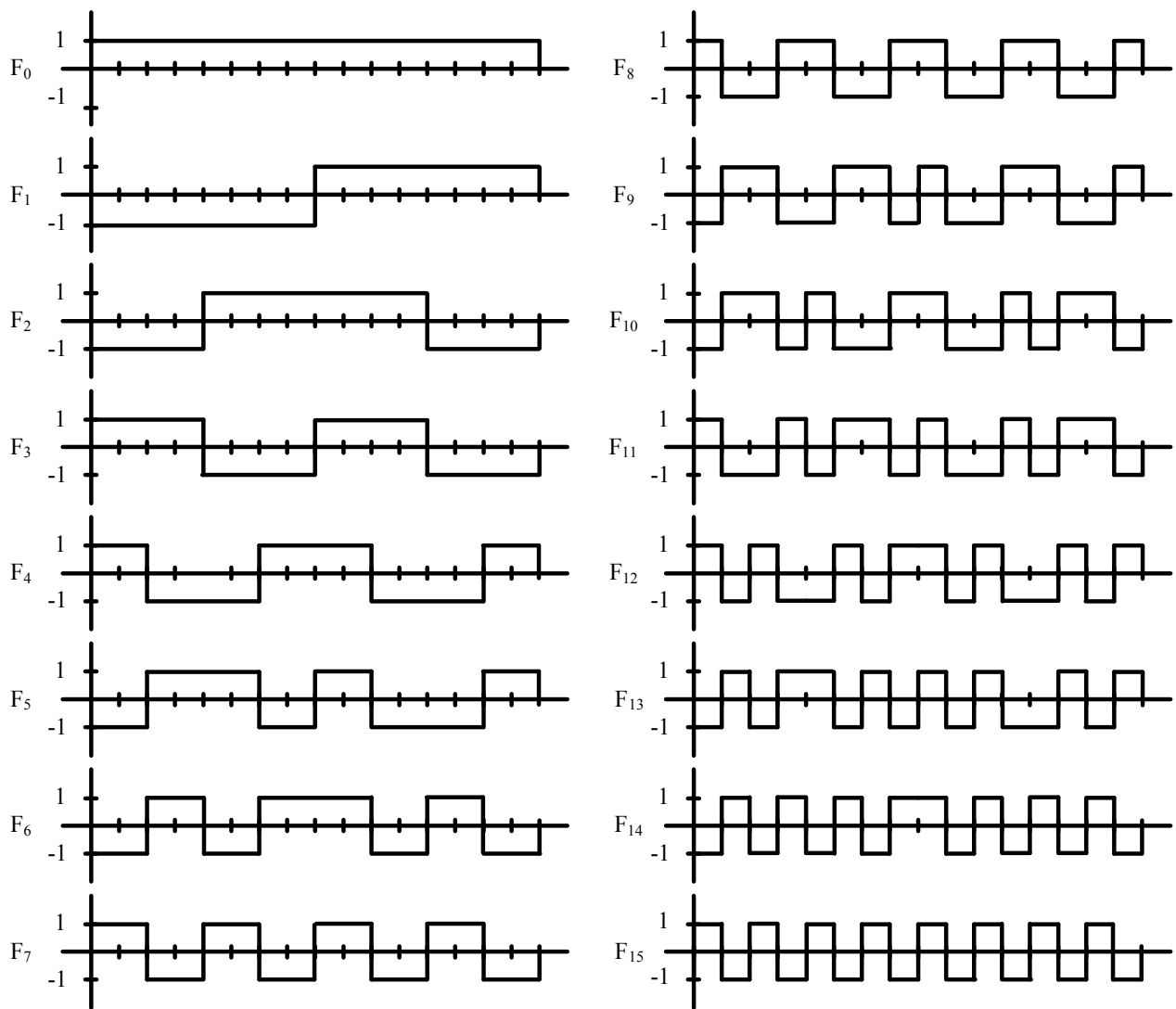


Рис. 4. Используемые фильтры

## 2. Огрубление спектрального представления

Цель выполнения огрубления спектрального представления – приведение значений спектральных коэффициентов спектра к конечному, заранее заданному, набору амплитуд. Это позволяет достичь инвариантности получаемого разложения к амплитуде спектральных коэффициентов. Входные данные алгоритма формирования огрублённого представления  $C$  – спектральное представление  $D$  и степень огрубления  $k$ .

Рассмотрим этапы выполнения алгоритма:

1) формирование множества чисел  $T$ :  $\{0, 1/k, 2/k, \dots, (k-1)/k, 1\}$  (например, для  $k = 3$  получим следующее множество:  $T = \{0, 0.33, 0.66, 1\}$ ) и множества промежутков  $W = \{(0; 1/k), [1/k; 2/k), \dots, [(k-1)/k, 1]\}$ , каждому из которых назначается номер  $N \in \overline{1, k}$ ;

2) обработка спектрального представления  $D = \{d_{ij}\}$ :

2.1) выбор максимального по абсолютной величине элемента спектра:

$$P = \max(d_{ij}), i = \overline{1, K}, j \in \overline{1, M_i};$$

2.2) умножение каждого элемента множества  $T$  на  $P$  (отклик нулевого фильтра не учитывается):

$$T_i = T_i \cdot P, i = \overline{1, |T|};$$

Допустим,  $P = 23$ , тогда  $T = \{0, 7.66, 15.33, 23\}$ ;

2.3) огрубление спектрального коэффициента  $d_{ij}(k)$ :

$$c_{ij}(k) = \begin{cases} \text{sign}(d_{ij}(k)) \cdot N, & \text{если } |d_{ij}(k)| \in (a_N, b_N); \\ 0, & \text{если } d_{ij}(k) = 0. \end{cases}$$

В приведённом правиле определяется промежуток  $((0; 7.66), [7.66, 15.33), [15.33, 23])$ , к которому относится  $k$ -ый спектральный коэффициент. Далее, в зависимости от полученного ответа, и с учётом знака коэффициента  $d_{ij}(k)$  рассматриваемому элементу  $c_{ij}(k)$  огрублённого спектра присваивается одно из значений: -3, -2, -1, 0, 1, 2, 3.

На рис. 5 показан пример огрубления спектрального разложения синуса на 4 уровне (частота – 33 Гц, длина сигнала – 4 периода,  $P = 488$ ).

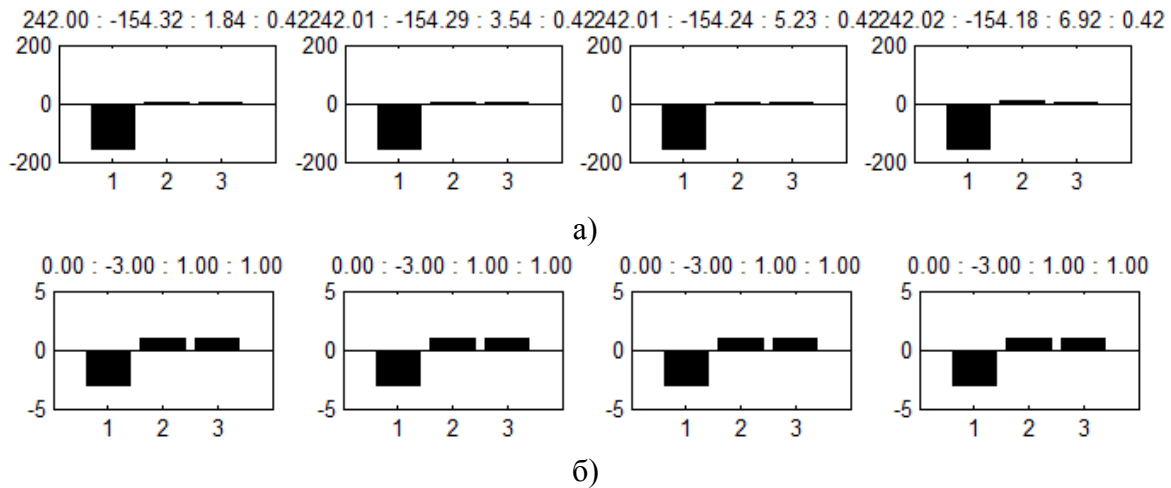
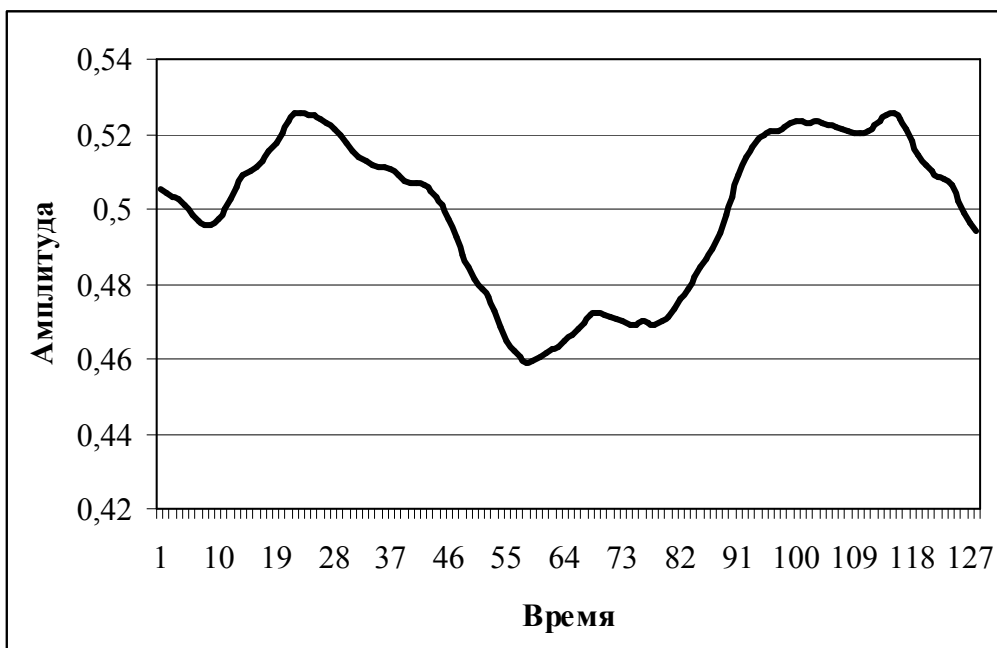


Рис. 5. Спектральное представление сигнала: а) исходное; б) огрублённое

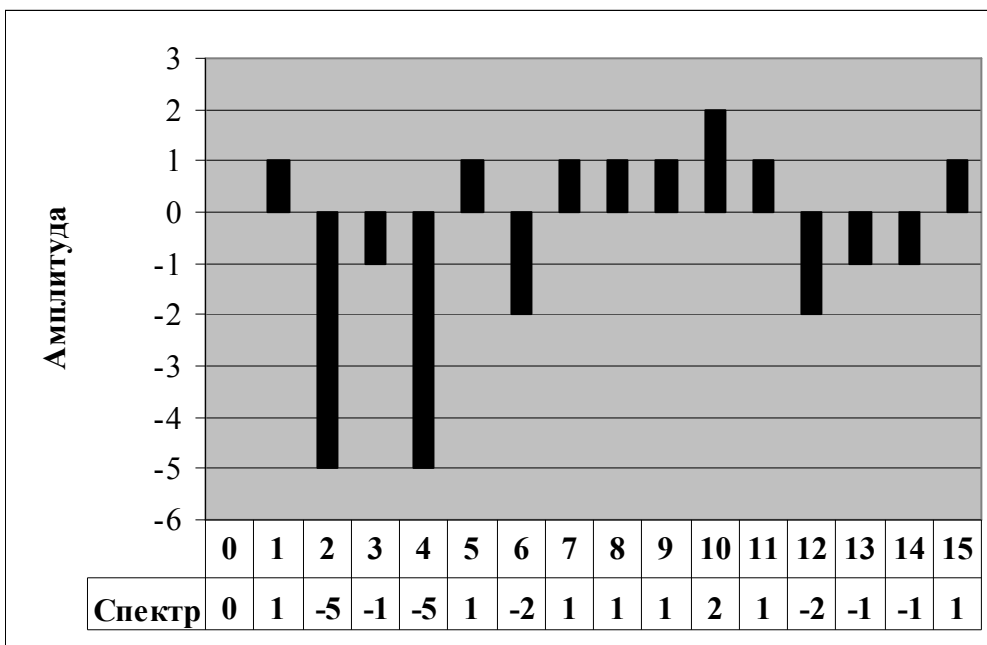
### 3. Определение ключевых сегментов

Разработанный алгоритм оценки частоты основного тона основан на использовании понятия «ключевой сегмент».

Ключевой сегмент соответствует участку сигнала, на котором максимален отклик небольшого количества операторов (одного - двух). Это означает, что сегмент сигнала подходит под образ операторов, давших максимальные отклики. На рис. 6 показан пример ключевого сегмента и вычисленного по нему спектра (нулевой коэффициент спектра на рисунке не отображается, при построении спектра использовалось 16 фильтров).



а)



б)

Рис. 6. Анализ сигнала: а) ключевой сегмент; б) спектр

Выбор ключевых сегментов сигнала  $f$  выполняется следующим образом:

- 1) на основе  $U$ -преобразования вычисляется спектральное представление  $D$  сигнала  $f$  (для получения представления сигнала используется только один уровень разложения);
- 2) формируется огрубленное представление  $C = \{c_{ij}\}$  спектрального представления  $D$ ;
- 3) анализируется каждый спектр  $c_{ij}$  огрубленного разложения  $C$ : если  $\forall t \in [1; L]$  верно, что  $|c_{ij}(t)| \leq k$  или  $|c_{ij}(t)| \geq l$  ( $k < l$ ), то сегмент сигнала  $f_{ij}$ , по которому вычислен спектр  $d_{ij}$ , считается ключевым.

Эксперимент показал, что анализ частоты повторения ключевых сегментов позволяет решить задачу оценки частоты основного тона сигнала.

Использование ключевых сегментов позволяет значительным образом (в среднем в 9 раз) уменьшить число анализируемых сегментов сигнала за счёт фильтрации сегментов, относящихся к шумовым участкам, паузам и т.д.

## 4. Алгебра групп

Алгебра групп используется для формирования описания звукового сигнала. С помощью замкнутых и полных групп выполняется спектрально-корреляционный анализ. Полные группы позволяют выявить корреляционные связи между операторами. Замкнутые группы позволяют выявить корреляционные связи между полными группами. Проводя аналогию между алгеброй групп и языком, можно ввести следующие соответствия: оператор – алфавит языка, полная группа – слово, замкнутая группа – словосочетание (см. рис. 7).

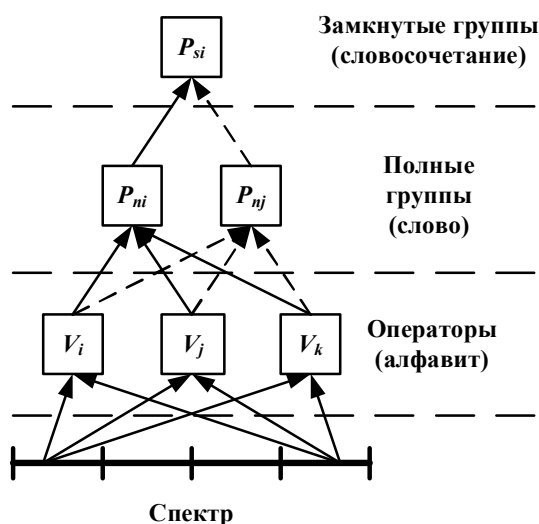


Рис. 7. Иерархия операторов и групп

Примем для множества фильтров  $\{F_i\}$ , что  $(+1 \rightarrow 1)$  и  $(-1 \rightarrow 0)$ . Получим множество бинарных операторов  $\{V_i\}$  (см. рис. 8). Для этих операторов допустимы теоретико-множественные операции объединения (сложения) и пересечения (умножения). Результатом имеем алгебру –  $A_V = \langle \{V_i\}; +, \times \rangle$  [9]. Необходимо отметить, что множества полных и замкнутых групп формируются на основе 16 операторов.

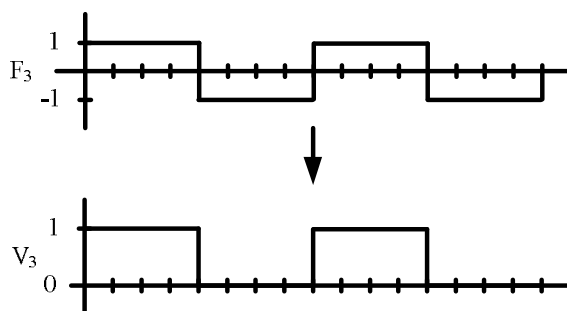


Рис. 8. Вычисление оператора по фильтру из 16 отсчётов

В алгебре  $A_V$  существуют алгебраические группы:

1)  $P_{ni}$  – группы на трёх элементах (названы полными), образованы на тройках операторов  $(V_i, V_j, V_k)$ , для которых справедливы соотношения:  $V_i + V_j + V_k \equiv e_1$  – единица. Полная группа допускает два описания – на операции пересечения  $V_i V_j V_k$  (операция умножения, число инверсий – чётно) и операции объединения  $\bar{V}_i + V_j + V_k$  (операции сложения, число инверсий – нечётно). Число возможных образов полных групп, с учётом инверсий операторов, равно четырём. Образ полной группы – компакт из четырех связанных элементов.

2)  $P_{si}$  – группы на четырёх элементах (названы замкнутыми), образованы на четвёрке операторов  $(V_i, V_j, V_n, V_m)$ , где  $(V_i, V_j, V_k) \in P_{ni}$ ,  $(V_n, V_m, V_k) \in P_{nj}$ , с описанием  $V_i V_j + V_n \bar{V}_m$  (где необходимое число инверсий операторов нечетно) и единицей  $-V_i + V_j + V_n + \bar{V}_m \equiv e_1$ . Число возможных образов замкнутых групп, с учётом инверсий операторов, равно восьми. Образ полной группы – компакт из восьми связанных элементов.

Множества  $\{P_{ni}\}$ ,  $\{P_{si}\}$  конечны и имеют мощность 35 и 105 соответственно. Множества  $\{V_i\}$ ,  $\{P_{ni}\}$ ,  $\{P_{si}\}$  – множества эталонов, используемые при решении задачи узнавания объекта исследования в пространстве эталонов. Обозначение полной группы на операции сложения –  $P_{nia}$ , на операции умножения –  $P_{nim}$ .

Одним из понятий алгебры групп в обработке изображений является понятие «визуальная масса» группы / изображения [9]. Применительно к звуковому сигналу будем использовать понятие «масса» группы / сигнала. Масса группы – сумма значений отсчётов сигнала, находящихся под образом группы.

Алгоритм вычисления массы группы:

$$\forall k \in \overline{1, L}$$

$$\text{if } B_{grp}[k] = 1$$

$$M_{grp} = M_{grp} + f[k]$$

где  $f[k]$  –  $k$ -ый отсчёт сигнала  $f$ ,  $B_{grp}$  – образ группы,  $M_{grp}$  – масса группы.

В первой строке табл. 2 приведён сигнал, состоящий из 16 отсчётов, во второй – образ полной группы  $\bar{V}_1 \bar{V}_8 V_9$ . Проводя вычисления на основе приведённого алгоритма, получаем что, масса группы составляет 3.6.

Таблица 3. Сигнал  $f$

0.1	0.9	1	0.3	0.2	1	0.7	0.3	0.3	0.1	0.2	0.1	0.1	0.1	0.1	0.1
0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0

#### 4.1. Полные группы

Полная группа формируется из трёх операторов  $V_i, V_j, V_k$  в том случае, если выполняется следующее условие:  $V_i + V_j + V_k = e_1$  ( $e_1 = V_0, i \neq j \neq k \neq 0$ ). Проанализировав все возможные сочетания операторов можно сгенерировать множество полных групп  $\{P_{ni}\}$ .

Рассмотрим операторы  $V_1, V_8$  и  $V_9$ . Результат сложения их образов доказывает, что эти операторы образуют полную группу (см. табл. 4).

Таблица 4. Проверка операторов

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	14
$V_1$	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
$V_8$	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1
$V_9$	0	1	1	0	0	1	1	0	1	0	0	1	1	0	0	1
$V_1 + V_8 + V_9$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Проверим наличие в спектре (см. табл. 5) группы  $V_1 V_8 V_9$ . Отклики операторов, составляющих группу, отличны от нуля, поэтому, группа  $V_1 V_8 V_9$  входит в спектр.

Значения операторов  $V_1$  и  $V_8$  отрицательны, следовательно, данные операторы являются инверсными. Число инверсных операторов чётно, поэтому, образ полной группы вычисляется на операции умножения (см. табл. 6, если число инверсных операторов нечётно, тогда образ полной группы вычисляется на операции сложения).

Таблица 5. Спектр сигнала



$V_0$	$V_1$	$V_2$	$V_3$	$V_4$	$V_5$	$V_6$	$V_7$	$V_8$	$V_9$	$V_{10}$	$V_{11}$	$V_{12}$	$V_{13}$	$V_{14}$	$V_{15}$
100	-12	-6	4	3	9	11	23	-2	5	21	11	3	0	7	8

Таблица 6. Вычисление образа группы  $\bar{V}_1\bar{V}_8V_9$

	$V_0$	$V_1$	$V_2$	$V_3$	$V_4$	$V_5$	$V_6$	$V_7$	$V_8$	$V_9$	$V_{10}$	$V_{11}$	$V_{12}$	$V_{13}$	$V_{14}$	$V_{15}$
$\bar{V}_1$	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
	×															
$\bar{V}_8$	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0
	×															
$V_9$	0	1	1	0	0	1	1	0	1	0	0	1	1	0	0	1
	=															
$\bar{V}_1\bar{V}_8V_9$	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0

Аналогичным образом можно проверить наличие в спектре любой другой полной группы.

## 4.2. Замкнутые группы

Замкнутую группу  $V_iV_jV_k\bar{V}_l$  можно описать тремя эквивалентными способами:  $V_iV_j + V_k\bar{V}_l$ ,  $V_iV_j + V_k\bar{V}_l$ ,  $V_i\bar{V}_j + V_kV_l$ .

Алгоритм построения замкнутой группы:

1) выбирается полная группа и формируется четыре её образа на операции умножения:  $V_iV_jV_k$ ,  $\bar{V}_i\bar{V}_jV_k$ ,  $\bar{V}_iV_j\bar{V}_k$ ,  $V_i\bar{V}_j\bar{V}_k$ ;

2) анализируются полученные образы: если в состав образа не входит оператор  $V_0$ , тогда операторы, входящие в него, образуют замкнутую группу.

Рассмотрим полную группу  $V_1V_8V_9$ . Варианты записи её образов приведены в табл. 7.

Таблица 7. Образы группы  $V_1V_8V_9$

	$V_0$	$V_1$	$V_2$	$V_3$	$V_4$	$V_5$	$V_6$	$V_7$	$V_8$	$V_9$	$V_{10}$	$V_{11}$	$V_{12}$	$V_{13}$	$V_{14}$	$V_{15}$
$V_1V_8V_9$	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	1
$\bar{V}_1\bar{V}_8V_9$	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0
$\bar{V}_1V_8\bar{V}_9$	1	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0
$V_1\bar{V}_8\bar{V}_9$	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0

Определим операторы, входящие в каждый образ:

- 1) образ группы  $V_1V_8V_9$  включает операторы  $V_8$ ,  $V_{11}$ ,  $V_{12}$  и  $V_{15}$ ;
- 2) образ группы  $\bar{V}_1\bar{V}_8V_9$  включает операторы  $V_1$ ,  $V_2$ ,  $V_5$  и  $V_6$ ;
- 3) образ группы  $\bar{V}_1V_8\bar{V}_9$  включает операторы  $V_0$ ,  $V_3$ ,  $V_4$  и  $V_7$ ; (по данной полной группе нельзя вычислить замкнутую);
- 4) образ группы  $V_1\bar{V}_8\bar{V}_9$  включает операторы  $V_9$ ,  $V_{10}$ ,  $V_{13}$  и  $V_{14}$ .

После обработки всех полных групп по приведённому алгоритму, генерируется множество замкнутых групп  $\{P_{si}\}$ .

Проверим наличие в спектре (см. табл. 5) группы  $V_8V_9V_{10}V_{11}$ . Значения операторов, составляющих группу, отличны от нуля, поэтому рассматриваемая группа входит в спектр.

Число инверсных операторов группы нечётно, следовательно, операторы  $V_8, V_9, V_{10}$  и  $V_{11}$  образуют замкнутую группу (если число инверсных операторов чётно, тогда они образуют замкнутое множество). Образ группы  $V_8V_9V_{10}V_{11}$  приведён в табл. 8.

Таблица 8. Вычисление образа группы  $V_8V_9V_{10}V_{11}$

	$V_0$	$V_1$	$V_2$	$V_3$	$V_4$	$V_5$	$V_6$	$V_7$	$V_8$	$V_9$	$V_{10}$	$V_{11}$	$V_{12}$	$V_{13}$	$V_{14}$	$V_{15}$
$\bar{V}_8$	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0
$V_9$	0	1	1	0	0	1	1	0	1	0	0	1	1	0	0	1
$V_{10}$	0	1	1	0	1	0	0	1	1	0	0	1	0	1	1	0
$V_{11}$	1	0	0	1	0	1	1	0	1	0	0	1	0	1	1	0
$\bar{V}_8V_9 + V_{10}V_{11}$	0	1	1	0	0	1	1	0	0	1	0	0	1	0	1	1

Аналогичным образом можно проверить наличие в спектре любой другой замкнутой группы.

## 5. Оценка частоты основного тона

Алгоритм оценки частоты основного тона на основе предлагаемого метода включает в себя следующие операции:

- 1) формирование набора ключевых сегментов  $K_f$ ;
- 2) извлечение данных из ключевых сегментов:
  - 2.1) формирование множества полных групп  $\{P_{ni}\}$  (см. пункт 4.1);
  - 2.2) формирование множества замкнутых групп  $\{P_{si}\}$  (см. пункт 4.2);
  - 2.3) вычисление по каждому ключевому сегменту  $K_{f,i}$  набора замкнутых групп  $\{P_{si,i}\}$ ,  $i \in \overline{1, N}$ , где  $N$  – количество ключевых сегментов;
  - 2.4) определение массы каждой группы набора  $\{P_{si,i}\}$ ;
  - 2.5) сохранение в наборе  $\{P_{si,i}\}$  первых  $N$  по массе групп;
- 3) сравнение наборов групп, сформированных по  $i$ -му и  $j$ -му сегментам ( $\{P_{si,i}\}$  и  $\{P_{sj,j}\}$ ): если состав групп совпадает не менее чем на  $M$  %, тогда считается, что сегменты сигнала подобны (с некоторой точностью) и расстояние  $dist$  между сегментами сохраняется в массив  $H$ :

$$H(dist) = H(dist) + 1;$$

- 4) определение номера максимального элемента  $X_{max}$  гистограммы  $H$ , который соответствует наиболее часто встречающемуся расстоянию между похожими спектрами (расстояние измеряется в отсчётах):

$$E_{max} = \max(H);$$

- 5) вычисление частоты основного тона:

$$P = \frac{F_d}{E_{max}},$$

где  $P$  – частота основного тона,  $F_d$  – частота дискретизации речевого сигнала.

## 6. Вычислительный эксперимент

Рассмотрим результаты оценки частоты основного тона на основе предлагаемого и существующих алгоритмов (см. табл. 3 – 7). Оценка качества вычисления частоты основного тона выполняется с помощью следующих показателей [1, 8]:

- 1) количество значительных ошибок оценки основного тона (gross pitch error):

$$G = \frac{1}{N_V} \sum_{i \in V} \delta_G(F_m(i), F_a(i)),$$

$$\delta_G(F_m(i), F_a(i)) = \begin{cases} 1, & |(F_m(i) - F_a(i)) / F_m(i)| > T, \\ 0, & \text{иначе;} \end{cases}$$

2) среднее значение незначительной ошибки оценки основного тона (fine pitch error):

$$S = \frac{1}{N_V} \sum_{i \in V} \delta_S(F_m(i), F_a(i)),$$

$$\delta_S(F_m(i), F_a(i)) = \begin{cases} |F_m(i) - F_a(i)|, & |(F_m(i) - F_a(i)) / F_m(i)| \leq T, \\ 0, & \text{иначе;} \end{cases}$$

3) среднеквадратическое отклонение незначительной ошибки оценки основного тона:

$$S_\sigma = \sqrt{\frac{1}{N_V} \sum_{i \in V} \delta_\sigma(F_m(i), F_a(i), S)},$$

$$\delta_\sigma(F_m(i), F_a(i), S) = \begin{cases} (F_m(i) - F_a(i))^2 - S^2, & |(F_m(i) - F_a(i)) / F_m(i)| \leq T, \\ 0, & \text{иначе.} \end{cases}$$

Используемые обозначения:  $F_m(i)$  – ручная оценка частоты основного тона  $i$ -го отсчёта сигнала;  $F_a(i)$  – автоматическая оценка частоты основного тона  $i$ -го отсчёта сигнала;  $T$  – порог, разделяющий грубые и незначительные ошибки оценки частоты основного тона;  $V$  – множество отсчётов исходного сигнала, относящихся к вокализованным сегментам;  $N_V$  – мощность множества  $V$ .

При тестировании алгоритмов использовалась база данных, описанная в [10]. База данных содержит записи голосов двух дикторов: мужчины (293.2 секунды) и женщины (370 секунд), ручную разметку записей на вокализованные / невокализованные сегменты, с указанием частоты основного тона для вокализованных сегментов.

Таблица 9. Сигнал без шума

Показатель/ Метод	АКФ [3]	YIN [4]	AMDF [5]	Кепстр [7]	YAAPT [8]	Предла- гаемый
$G$	0.0313	0.0309	0.0753	0.1636	0.1685	0.0498
$S$	0.0472	0.0245	0.0063	0.0381	0.0227	0.0250
$S_\sigma$	0.0461	0.0277	0.0111	0.0337	0.0237	0.0284

Таблица 10. Сигнал искажён аддитивным равномерным шумом, ОСШ = 10 дБ

Показатель/ Метод	АКФ	YIN	AMDF	Кепстр	YAAPT	Предла- гаемый
$G$	0.1519	0.0759	0.7567	0.3444	0.2462	0.1122
$S$	0.0489	0.0297	0.0202	0.0576	0.0225	0.0309
$S_\sigma$	0.0385	0.0294	0.0234	0.0521	0.0219	0.0360

Таблица 11. Сигнал искажён аддитивным равномерным шумом, ОСШ = 0 дБ

Показатель/ Метод	АКФ	YIN	AMDF	Кепстр	YAAPT	Предла- гаемый
$G$	0.6035	0.5033	0.9478	0.6057	0.5678	0.5760
$S$	0.0721	0.0414	0.0396	0.0839	0.0265	0.0521
$S_\sigma$	0.0578	0.0436	0.0456	0.0603	0.0265	0.0429

Таблица 12. Сигнал искажён аддитивным нормальным шумом, ОСШ = 10 дБ

Показатель/ Метод	АКФ	YIN	AMDF	Кепстр	УААРТ	Предла- гаемый
$G$	0.3871	0.2151	0.8342	0.4258	0.2812	0.3455
$S$	0.0400	0.0347	0.0416	0.0509	0.0206	0.0502
$S_{\sigma}$	0.0378	0.0361	0.0517	0.0495	0.0164	0.0394

Таблица 13. Сигнал искажён аддитивным нормальным шумом, ОСШ = 0 дБ

Показатель/ Метод	АКФ	YIN	AMDF	Кепстр	УААРТ	Предла- гаемый
$G$	0.6792	0.7646	0.9752	0.5585	0.8243	0.7839
$S$	0.0634	0.0643	0.0815	0.0919	0.0246	0.0580
$S_{\sigma}$	0.0519	0.0564	0.0391	0.0564	0.0252	0.0446

## Заключение

В работе рассматривается алгоритм оценки частоты основного тона голоса, основанный на теории активного восприятия. Анализ результатов вычислительного эксперимента позволяет сделать вывод, что разработанный алгоритм успешно справляется с поставленной задачей оценки частоты основного тона по неискажённым и искажённым записям голоса человека.

По сравнению с другими алгоритмами, работающими во временной области, предлагаемый алгоритм позволяет обрабатывать сложные периодические сигналы (с большим количеством гармоник) без предварительного сглаживания сигнала фильтрами нижних частот.

Результаты, полученные в работе, могут использоваться при разработке алгоритмов сжатия, улучшения качества звукового сигнала, настройке музыкальных инструментов, при создании нотной записи и классификации музыкальных произведений.

## Литература

1. L. R. Rabiner, M. J. Cheng, A. E. Rosenberg, C. A. McGonegal A Comparative performance study of several pitch detection algorithms // IEEE Transactions on acoustics, speech, and signal processing, V. ASSP-24, N. 5. October 1976, P. 399-418.
2. B. Kedem Spectral analysis and discrimination by zero-crossings. Proceedings of the IEEE, 74(11):1477–1493, November 1986.
3. А. В. Аграновский, Д. А. Леднов Теоретические аспекты алгоритмов обработки и классификации речевых сигналов Москва: Радио и связь, 2004. – 164 с.
4. A. Cheveigne, H. Kawahara Yin, a fundamental frequency estimator for speech and music. Journal of the Acoustical Society of America, 111(4), 2002.
5. M. J. Ross Average magnitude difference function pitch extractor // IEEE Transactions on Speech and Audio processing, V. 22, 1974, P. 353-362.
6. A. M. Noll Pitch determination of human speech by the harmonic product spectrum; the harmonic sum spectrum and a maximum likelihood estimate // Proceedings of the Symposium on Computer Processing in Communications, Polytech. Inst. of Brooklyn, 1969, pp. 779-797.
7. F. Wang, P. Yip Cepstrum analysis using discrete trigonometric transforms // IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 39(2), 1991, P. 538-541.
8. S. A. Zahorian, H. Hu A spectral/temporal method for robust fundamental frequency tracking // Journal of Acoustic Society of America, N. 123 (6), June 2008, P. 4559-4571.
9. В. А. Утробин Физические интерпретации элементов алгебры изображения // Успехи физических наук, Т. 174, № 10, 2004, С. 1089–1104.

10. P. C. Bagshaw, S. M. Hiller, M. A. Jack Enhanced pitch tracking and the processing of F0 contours for computer and intonation teaching // Proceedings of European Conference on Speech Communication, 1993, P. 1003-1006.