

МЕТОДИКА ОЧИСТКИ ЗАШУМЛЕННОГО СИГНАЛА, УЧИТЫВАЮЩАЯ КЛАССИФИКАЦИЮ ЗВУКОВ РЕЧИ

Федоров Е.Е.

Донецкий национальный университет, г.Донецк
кафедра прикладной математики и теории систем управления
E-mail: fee@iai.donetsk.ua

Abstract

Fedorov U. Technique of cleaning of a noisy signal taking into account classification of sounds of speech. At speech recognition the relevant role is played by preliminary clearing of a signal of a noise. For the solution of this problem in the article the technique of clearing of a noisy signal taking into account classification of notes of speech is offered. This technique will use a two-channel filtration, founded on discrete wavelet-transformation.

Общая постановка проблемы. В современной отечественной и мировой практике активно ведутся разработки естественно-языковых систем общения человек–компьютер, одной из составных частей которых может быть система распознавания речи. При создании системы распознавания необходимо решить задачу очистки зашумленного входного сигнала.

Анализ исследований. В работах [1–3] приводится методика определения границ речи, учитывающая классификацию звуков речи [4], но не подавление шума. В работах [5–7], рассматриваются методы очистки сигнала, не учитывающие классификацию звуков речи.

Постановка задачи исследования. Разработать методику очистки сигнала от шума, учитывающую классификацию звуков речи.

Решение задачи. В статье рассматривается схема двухканальной фильтрации, базирующаяся на дискретном вейвлет-преобразовании; предлагается авторская методика ФЕОС очистки сигнала от шума; учитывающая классификацию звуков речи; приводятся результаты работы программы, основанной на этой методике.

Перед рассмотрением авторской методики очистки зашумленного сигнала, опишем используемую ей **схему двухканальной фильтрации**, основанной на дискретном вейвлет-преобразовании. Входной сигнал $x(n)$ разбит на сегменты $x_s(n)$, которые представим в виде c_{s0n} . Тогда спектр C_{s0} представлен в виде:

$$C_{s0}(e^{-j\omega_k \Delta t}) = \sum_{n=0}^{N-1} c_{s0n} e^{-j\omega_k n \Delta t}, \quad k \in \overline{0, N-1}, \quad s \in \overline{1, S}, \quad (1)$$

где N — длина сегмента сигнала, $N = 2^d$

Передаточная функция (частотная характеристика) КИХ-фильтра высоких частот для анализа и синтеза

$$G(e^{-j\omega_k \Delta t}) = \sum_{l=0}^{L_G-1} g_l e^{-j\omega_k l \Delta t}, \quad E(e^{-j\omega_k \Delta t}) = \sum_{l=0}^{L_E-1} e_l e^{-j\omega_k l \Delta t}, \quad (2)$$

где L_G, L_E — длины фильтров, Δt — шаг квантования.

Передаточная функция КИХ-фильтра нижних частот для анализа и синтеза

$$H(e^{-j\omega_k \Delta t}) = \sum_{l=0}^{L_H-1} h_l e^{-j\omega_k l \Delta t}, \quad F(e^{-j\omega_k \Delta t}) = \sum_{l=0}^{L_F-1} f_l e^{-j\omega_k l \Delta t}, \quad (3)$$

где L_H, L_F — длины фильтров.

Если в качестве фильтра выбран вейвлет Добеши (рис. 1),

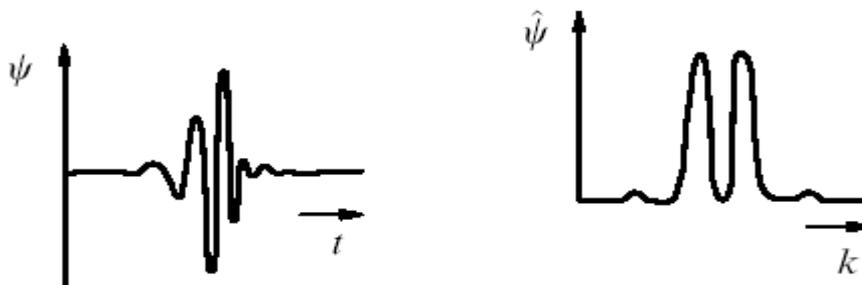


Рисунок 1 — Вейвлеты Добеши во временной и частотной областях для $M=4$

$$\psi(t) = \sqrt{2} \sum_{l=0}^{2M-1} (-1)^l h_{(2M-1)-l} \varphi(2t-l), \quad \hat{\psi}(k) = \sqrt{2} \sum_{l=0}^{2M-1} h_l \varphi(2t-l), \quad (4)$$

где $\varphi(t)$ — масштабирующая функция, M — порядок вейвлета (число нулевых моментов), то набор свойств масштабирующих функций и вейвлетов, обеспечивающих ортогональность масштабирующих функций, ортогональность вейвлетов масштабирующим функциям, ортогональность вейвлетов полиномам до степени $M-1$, условие нормировки, сводится к системе (5), которая однозначно определяет коэффициенты h_l и g_l :

$$\begin{cases} \sum_{l=0}^{2M-1} h_l h_{l+2m} = \delta_{0m} \\ g_l = (-1)^l h_{2M-l-1}, l \in \overline{0, M-1} \\ \sum_{l=0}^{2M-1} (-1)^l l^p h_l = 0, p \in \overline{0, M-1} \\ \sum_{l=0}^{2M-1} h_l = \sqrt{2} \end{cases} \quad (5)$$

Сигнал подвергается преобразованию в системе анализа-синтеза

Анализ (декомпозиция) сигнала:

Для каждого сегмента $s, s \in \overline{1, S}$, производится декомпозиция на P уровней в два этапа. При этом на текущем i -том уровне, $i \in \overline{1, P}$, выполняется:

а) свертка сигнала с полосовыми фильтрами для i -го уровня декомпозиции сигнала для получения высоко- и низкочастотных составляющих (субполос) представлена в виде

$$D_{si}(e^{-j\omega_k \Delta t}) = a_0^{1/2} G(e^{-j\omega_k \Delta t}) C_{s,i-1}(e^{-j\omega_k \Delta t}), \quad k \in \overline{0, N/2^{i-1}-1}, \quad (6)$$

$$C_{si}(e^{-j\omega_k \Delta t}) = a_0^{1/2} H(e^{-j\omega_k \Delta t}) C_{s,i-1}(e^{-j\omega_k \Delta t}), \quad k \in \overline{0, N/2^{i-1}-1}. \quad (7)$$

б) уменьшение числа высоко- и низкочастотных составляющих D_{si} и C_{si} вдвое

$$\hat{D}_{si}(e^{-j\omega_k \Delta t}) = D_{si}(e^{-j2\omega_k \Delta t}), \quad k \in \overline{0, N/2^i-1}, \quad (8)$$

$$\hat{C}_{si}(e^{-j\omega_k \Delta t}) = C_{si}(e^{-j2\omega_k \Delta t}), \quad k \in \overline{0, N/2^i-1}. \quad (9)$$

в) Применение пункта а) к низкочастотной части $\hat{C}_{si}, i := i + 1$.

Синтез (реконструкция) сигнала:

Реконструкция каждого сегмента s , $s \in \overline{1, S}$, на основе соответствующих коэффициентов декомпозиции по P уровням производится в два этапа. При этом на текущем i -том уровне, $i \in \overline{1, P}$, выполняется:

а) Увеличение числа высоко- и низкочастотных составляющих \hat{D}_{si} и \hat{C}_{si} вдвое путем вставки нулей между элементами

$$D_{si}(e^{-j2\omega_k \Delta t}) = \hat{D}_{si}(e^{-j\omega_k \Delta t}), D_{si}(e^{-j\omega_k \Delta t}) = 0, k \in \overline{0, N/2^i - 1}, \quad (10)$$

$$C_{si}(e^{-j2\omega_k \Delta t}) = \hat{C}_{si}(e^{-j\omega_k \Delta t}), C_{si}(e^{-j\omega_k \Delta t}) = 0, k \in \overline{0, N/2^i - 1}. \quad (11)$$

б) Свертка высоко- и низкочастотных составляющих сигнала с полосовыми фильтрами с последующим сложением двух сверток представлена в виде

$$C_{s,i-1}(e^{-j\omega_k \Delta t}) = a_0^{1/2} (F_i(e^{-j\omega_k \Delta t})C_{si}(e^{-j\omega_k \Delta t}) + G_i(e^{-j\omega_k \Delta t})D_{si}(e^{-j\omega_k \Delta t})), \quad (12)$$

$$k \in \overline{0, N/2^i - 1}.$$

в) $i := i - 1$, переход к пункту а)

Результатом синтеза являются сегменты сигнала $C_{s0}(e^{-j\omega_k \Delta t})$.

Отметим, что константу $a_0^{1/2} = \sqrt{2}$ выносят из блоков анализа-синтеза.

Обобщенно процесс декомпозиции-реконструкции (или анализа-синтеза) представлен на рис.2

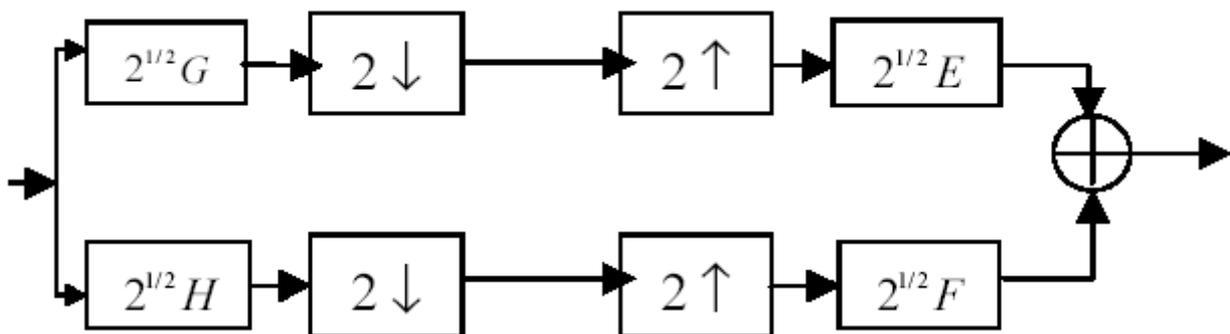


Рисунок 2 — Схема двухполосного банка фильтров

Разработанная автором **методика ФЕОС очистки зашумленного сигнала** использует приведенную выше двухканальную фильтрацию и заключается в следующем.

1) Производится декомпозиция каждого s -го сегмента на P уровней согласно (6)–(9)

2) Для s -го сегмента, $s \in \overline{1, S}$, по формулам (13), (14) вычисляются энергии высокочастотных составляющих на каждом i -м уровне, $i \in \overline{1, P}$, энергия низкочастотных составляющих — только на P -м уровне:

$$DE_{si} = \sum_k \hat{D}_{si}^2(e^{-j\omega_k \Delta t}), k \in \overline{0, N/2^{i-1} - 1}, \quad (13)$$

$$CE_s = \sum_k \hat{C}_{sP}^2(e^{-j\omega_k \Delta t}), k \in \overline{0, N/2^{i-1} - 1}. \quad (14)$$

3) Поскольку удаление шума из сигнала заключается в обнулении детализирующих и аппроксимирующих коэффициентов, которые меньше максимальных шумовых детализирующих и аппроксимирующих коэффициентов, производится расчет максимальных коэффициентов, состоящий в следующем.

На совокупности s -х сегментов, которые предшествуют началу речи, т.е. удовлетворяют условию (15),

$$\forall i \in \overline{1, P} \quad DE_{si} < \alpha, \text{ где } \alpha \text{ — порог,} \quad (15)$$

определяются максимальные значения энергии для высокочастотных (16) и низкочастотных (17) составляющих

$$MaxDE_i = \max_s DE_{si}, \quad i \in \overline{1, P}, \quad (16)$$

$$MaxCE = \max_s CE_s, \quad (17)$$

и номера соответствующих сегментов (18) и (19)

$$ArgMaxDE_i = \arg \max_s DE_{si}, \quad i \in \overline{1, P}, \quad (18)$$

$$ArgMaxCE = \arg \max_s CE_s. \quad (19)$$

Для уровней сегментов с максимальными значениями энергий находятся максимальные детализирующие (20) и аппроксимирующие (21) коэффициенты

$$MaxD_i = \max_k \hat{D}_{si}(e^{-j\omega_k \Delta t}), \text{ где } s = ArgMaxDE_i, \quad i \in \overline{1, P}, \quad (20)$$

$$MaxC = \max_k \hat{C}_{sP}(e^{-j\omega_k \Delta t}), \text{ где } s = ArgMaxCE. \quad (21)$$

4) Производится обнуление детализирующих и аппроксимирующих коэффициентов в соответствии с типом сегмента:

а) Если сегмент содержит только шум, т.е. выполняется условие (22),

$$\forall i \in \overline{1, P} \quad DE_{s,i} < \alpha, \text{ где } \alpha \text{ — порог,} \quad (22)$$

обнуляются все детализирующие (23) и аппроксимирующие (24) коэффициенты

$$\hat{D}_{si}(e^{-j\omega_k \Delta t}) = 0, \quad i \in \overline{1, P}, \quad k \in \overline{0, N/2^{i-1} - 1}, \quad (23)$$

$$\hat{C}_{sP}(e^{-j\omega_k \Delta t}) = 0, \quad k \in \overline{0, N/2^{i-1} - 1}. \quad (24)$$

б) Если сегмент содержит шумные глухие щелевые и смычно-щелевые звуки, т.е. выполняется условие (25)

$$DE_{s,1} > \beta, \text{ где } \beta \text{ — порог,} \quad (25)$$

и если для аппроксимирующих коэффициентов выполняется условие (26)

$$\hat{C}_{sP}(e^{-j\omega_k \Delta t}) < MaxC, \quad (26)$$

то они обнуляются

$$\hat{C}_{sP}(e^{-j\omega_k \Delta t}) = 0. \quad (27)$$

в) Если сегмент содержит тональные (шумные звонкие, сонорные или гласные) звуки, т.е. не выполняется условие (25), то:

– если для детализирующих коэффициентов выполняется условие (28)

$$\hat{D}_{si}(e^{-j\omega_k \Delta t}) < MaxD_i, \quad (28)$$

то они обнуляются

$$\hat{D}_{si}(e^{-j\omega_k \Delta t}) = 0; \quad (29)$$

– если для детализирующих коэффициентов выполняется условие (30)

$$\hat{C}_{sP}(e^{-j\omega_k \Delta t}) < MaxC, \quad (30)$$

то они обнуляются

$$\hat{C}_{sP}(e^{-j\omega_k \Delta t}) = 0. \quad (31)$$

5) По полученным новым коэффициентам декомпозиции производится реконструкция сигнала согласно (10)–(12)

Алгоритм очистки зашумленного сигнала представлен на рис.3. Эта блок-схема функционирует следующим образом:

шаг 1 — задание параметров α, β, N, M

шаг 2 — расчет $\hat{D}_{si}(e^{-j\omega_k \Delta t})$ и $\hat{C}_{sP}(e^{-j\omega_k \Delta t})$ согласно (6)–(9)

шаг 3 — расчет DE_{si} и CE_s согласно (13)–(14)

шаг 4 — $\forall i \in \overline{1, P} \ DE_{si} < \alpha$

шаг 5 — расчет $MaxD_i$ и $MaxC$ согласно (16)–(21)

шаг 6 — $\forall i \in \overline{1, P} \ DE_{si} < \alpha$

шаг 7 — обнуление $\hat{D}_{si}(e^{-j\omega_k \Delta t})$ и $\hat{C}_{sP}(e^{-j\omega_k \Delta t})$ согласно (23)–(24)

шаг 8 — $DE_{s,1} > \beta$

шаг 9 — $\hat{C}_{sP}(e^{-j\omega_k \Delta t}) < MaxC$

шаг 10 — обнуление $\hat{C}_{sP}(e^{-j\omega_k \Delta t})$ согласно (27)

шаг 11 — $\hat{D}_{si}(e^{-j\omega_k \Delta t}) < MaxD_i$

шаг 12 — обнуление $\hat{D}_{si}(e^{-j\omega_k \Delta t})$ согласно (29)

шаг 13 — $\hat{C}_{sP}(e^{-j\omega_k \Delta t}) < MaxC$

шаг 14 — обнуление $\hat{C}_{sP}(e^{-j\omega_k \Delta t})$ согласно (31)

шаг 15 — расчет $C_{s,i-1}(e^{-j\omega_k \Delta t})$.

В соответствии с предложенной методикой, была *разработана программа и получены результаты ее работы*.

В качестве длины сегмента выбиралось $N=512$ (примерно 20 мс), при этом сегменты не перекрывались. В качестве фильтра-вейвлета выбирался вейвлет Добеши порядка $M=4$, т.е. $L_G = L_E = L_H = L_F = 8$. Количество уровней $P=5$. На нижеприведенных рисунках можно видеть результаты работы программы для слова «Саша»: на рис.4 представлен зашумленный сигнал, на рис.5 — очищенный сигнал.

Выводы.

1) В статье была предложена методика ФЕОС очистки зашумленного сигнала, учитывающая классификацию звуков речи. Эта методика использует двухканальную фильтрацию, основанную на дискретном вейвлет-преобразовании. На основании этой методики создана программа, результаты работы которой приведены в статье.

2) Основные положения данной работы предназначены для реализации в интеллектуальных системах управления, в которых команды поступают на естественном языке.

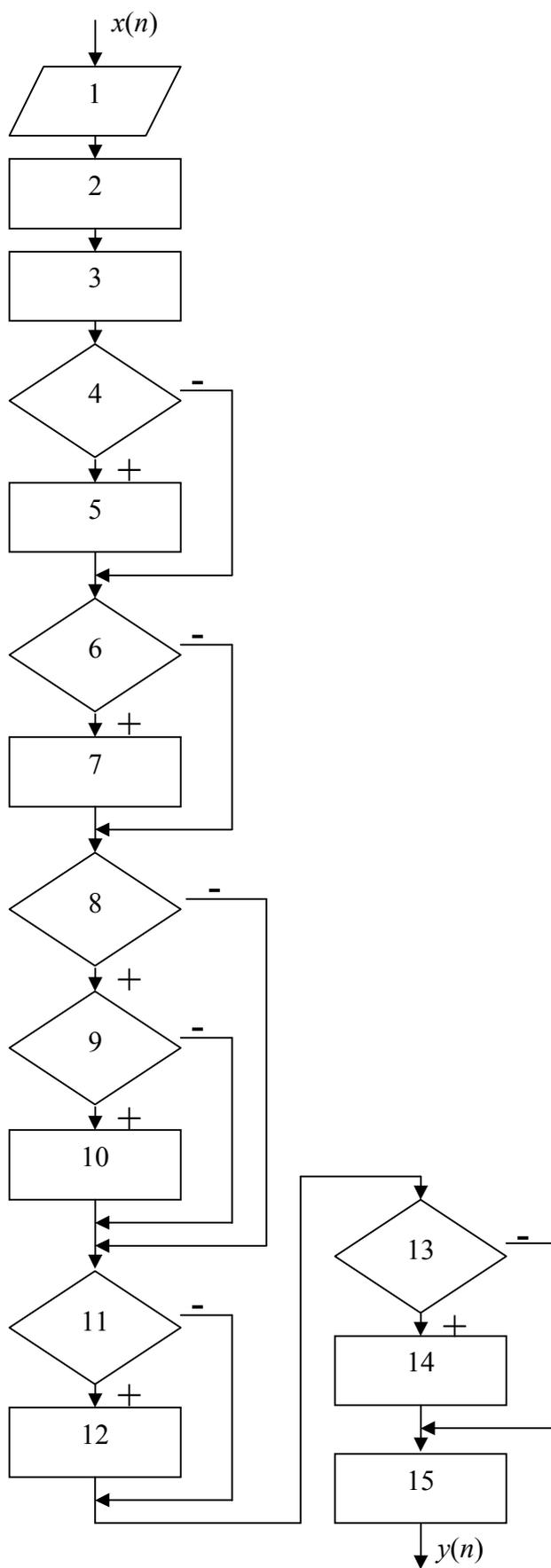


Рисунок 3 — Блок-схема методики очистки зашумленого сигнала

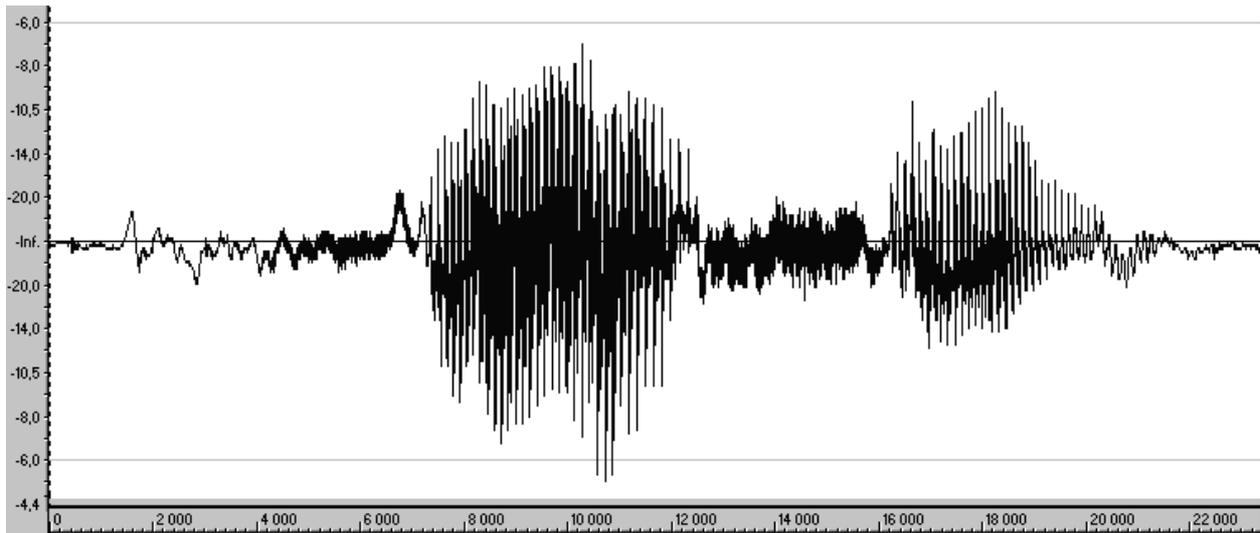


Рисунок 4 — Зашумленный сигнал

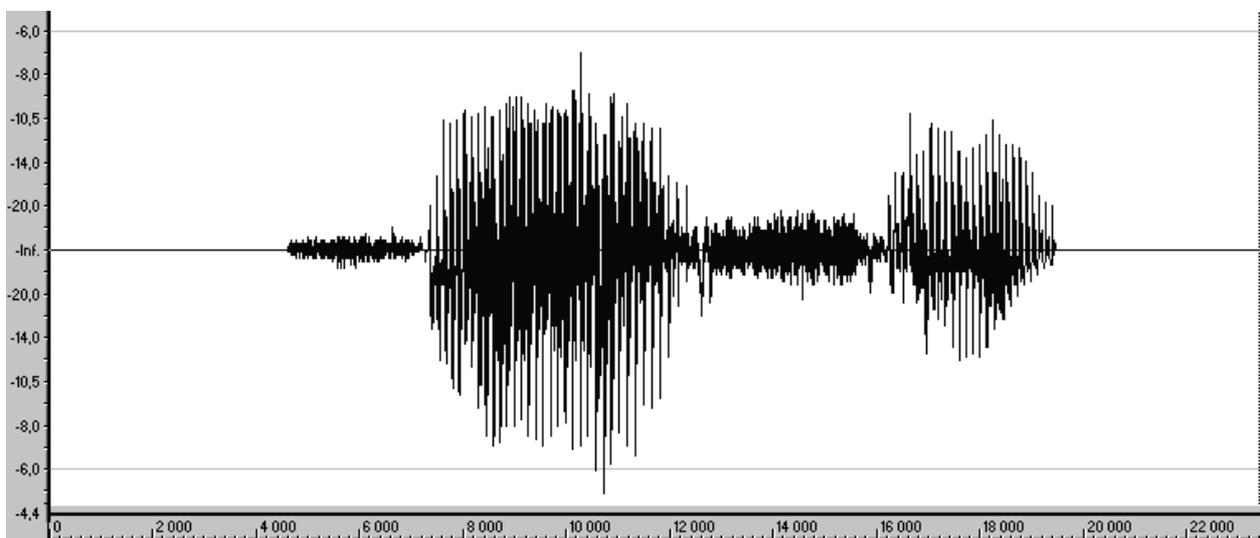


Рисунок 5 — Очищенный сигнал

Литература

1. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов. — М.: Радио и связь, 1981. — 496 с.
2. Федоров Е.Е., Шелепов В.Ю. Автоматическое определение начала и конца записи речи // Искусственный интеллект. — 2002. — № 4. — С. 295–298.
3. Ермоленко Т. В. Фонетический анализ речевого сигнала на основе вейвлет-разложения // Искусственный интеллект. — 2003. — №3. — С. 409–416.
4. Современный русский язык: Учеб. для филол. спец. высших учебных заведений / Под ред. В.А.Белошапковой. — М.: Азбуковник, 1997. — 928 с.
5. Секунов Н.Ю. Обработка звука на РС. — СПб.: БХВ – Санкт-Петербург, 2001. — 1248 с.
6. Freeman D., Sonthcott C., Boyd I.A. Voice activity detector for the Pan-European digital cellular mobile telephone service // IEE Colloquium «Digitized Speech Communication via Mobile Radio». — London. — 1988. — P. 61–65.
7. Дьяконов В.П. Вейвлеты. От теории к практике. — М.: СОЛОН-Р, 2002. — 448 с.