УДК 004.934: 004.42: 61:00.7

Д.А. Вольф, Р.В. Мещеряков

МОДЕЛЬ И ПРОГРАММНАЯ РЕАЛИЗАЦИЯ СИНГУЛЯРНОГО ОЦЕНИВАНИЯ ЧАСТОТЫ ОСНОВНОГО ТОНА РЕЧЕВОГО СИГНАЛА

Вольф Д.А., Мещеряков Р.В. **Модель и программная реализация сингулярного оценивания частоты основного тона речевого сигнала.**

Аннотация. В статье рассматривается сингулярная модель оценивания частоты основного тона речевого сигнала, а также ее программная реализация. Применение модели сингулярного оценивания частоты основного тона позволяет уменьшить вычислительную сложность алгоритмов анализа речевого сигнала путем аппроксимации края сингулярного спектра и обеспечить меньшее количество ошибок оценивания частоты основного тона за счет использования сингулярной модели вокализированного сегмента речи, учитывающей нестационарные параметры основного тона с помощью собственных чисел. Программная реализация модели используется в модуле расчетов комплекса программ речевой реабилитации онкологических больных после резекции гортани.

Ключевые слова: оценивание частоты основного тона речевого сигнала, сингулярный спектральный анализ речи, модель, программная реализация.

Volf D.A., Meshcheryakov R. V. **Software Implementation of a Singular Meter of the Pitch Frequency of a Speech Signal.**

Abstract. The article deals with software implementation of the evaluation of the pitch frequency of the speech signal based on the mathematical apparatus for singular spectral analysis. The program is used in calculation module of a program complex for speech rehabilitation of cancer patients after resection of larynx used in rehabilitation training of patients after complete or partial loss of sounding speech as a result of laryngectomy.

Keywords: estimation of the pitch frequency of the speech signal; singular spectrum analysis of speech; model; software implementation.

1. Введение. Поставленная в работе задача определения частоты основного тона речевого сигнала, включая распределение амплитуд, периодов и начальных фаз гармоник, образующих сложный полигармонический сигнал, остается все еще нерешенной и активно исследуемой в области речевых технологий [1–3]. Существующие алгоритмы оценивания ЧОТ [4–7] позволяют проводить анализ статистических данных без учета особенностей речеобразования и речевосприятия, связанных с анатомией и физиологией человека, так как методы анализа [8], лежащие в их основе, ограничены периодической (стационарной) моделью речевого сигнала, которая подразумевает точное повторение периода и амплитуды основного тона и не допускает их изменения на протяжении окна анализа. В свою очередь, это влияет на точность результатов оценивания ЧОТ [9].

Исходя из данной проблемы, появляется мотивация к разработке такой модели, которая позволит осуществлять учет нестационарных

амплитуд, периодов и фаз гармоник, входящих в речевой сигнал. С другой стороны, повышение точности вычисления ЧОТ приводит к увеличению вычислительной сложности [10]. Таким образом, разработка новых методов анализа речи для задач оценивания частоты основного тона речи, является актуальной [11].

Целью настоящей работы является получение модели оценивания частоты основного тона речевого сигнала при оптимальной временной обработке с учетом особенностей речеобразования и речевосприятия, связанных с анатомией и физиологией человека, а также получения ее программной реализации. Новизна данной работы заключается в применении математического аппарата сингулярного спектрального анализа к обработке речевых сигналов.

В рамках базовой части государственного задания ТУСУР (проект № 3657 от 2015г.) для НИИ Онкологии г. Томска разработан программный комплекс речевой реабилитации онкологических больных после резекции гортани (Свидетельство о государственной регистрации программы для ЭВМ № 2015618857 – "Программа речевой реабилитации больных после резекции гортани"). Разработанный программный комплекс состоит из семи модулей каждый из которых представляет собой черный ящик, который принимает на вход речевые данные, обрабатывает их и возвращает обратно интерпретируемый результат. Одним из ключевых модулей является модуль расчетов, в котором решаются задачи вычисления параметров речи. Одной из решаемых задач является оценивание частоты основного тона (ЧОТ) речевого сигнала.

2. Сингулярная модель вокализированного сегмента речи и сингулярная модель оценивания частоты основного тона. Особенность предлагаемого метода оценивания ЧОТ заключается в разложение речевого сигнала в элементарный спектр временных рядов (квазигармоник) посредством сингулярного спектрального анализа с последующим выбором квазигармонической составляющей, соответствующей основному тону речи. Рассмотрим прямую задачу. Пусть ряд S_N, полученный в результате процедуры дискретизации речевого сигнала $S(t)$, принимается в качестве фонемного ряда. С фонемным рядом проводится процедура Ганкелизации [12, 13]:

$$\mathbf{A}=[S_{i-1},\ldots,S_{i+L-1}]^{\mathrm{T}},\ 1\le i\le K,\ K = N - L + 1, \qquad (1)$$

таким образом получается траекторная матрица **A**, состоящая из K векторов вложений длины L. Для траекторной матрицы (1) вычисляется матричное разложение вида:

$$\mathbf{A} = \sum_{i=0}^{L\text{-}1} \mathbf{A}^{<i>} = \sum_{i=0}^{L\text{-}1} (\sqrt{\lambda_i}\,\mathbf{u}^{<i>})[\mathbf{x}^{<i>}]^{\mathrm{T}}, \qquad (2)$$

где: λ_i – i-е собственное значение ковариационной матрицы $\mathbf{A}\mathbf{A}^{\mathrm{T}}$;
$\mathbf{u}^{<i>}$ – i-й собственный вектор ковариационной матрицы $\mathbf{A}\mathbf{A}^{\mathrm{T}}$;
$\mathbf{x}^{<i>}$ – i-й собственный вектор (главных компонент), образованный строками матрицы **A**.

Для произведения векторов в (2) вычисляется матричное усреднение по диагонали:

$$\mathbf{T}_j^{<n>} = \begin{cases} \dfrac{1}{j+1} \displaystyle\sum_{i=0}^{j} [\sqrt{\lambda_n}\,\mathbf{u}^{<n>}\mathbf{x}^{<n>\mathrm{T}}]_{iK+j-i}, 0 \le j < L; \\ \dfrac{1}{L} \displaystyle\sum_{i=0}^{L-1} [\sqrt{\lambda_n}\,\mathbf{u}^{<n>}\mathbf{x}^{<n>\mathrm{T}}]_{iK+j-i}, L \le j < K; \\ \dfrac{1}{N-j} \displaystyle\sum_{i=0}^{L-1-(j-K)} [\sqrt{\lambda_n}\,\mathbf{u}^{<n>}\mathbf{x}^{<n>\mathrm{T}}]_{(j-K+i)K+K-1-i}, K \le j < N. \end{cases} \qquad (3)$$

в результате которого образуется матрица временных рядов $\mathbf{T}_{L,N}$ в строках которого содержится квазигармонический спектр. Аналогично тому, как в гармонических моделях осуществляется проекция речевого сигнала в гармонический базис (например, в преобразованиях Фурье) [14, 15], так и в прямой задаче сингулярного спектрального анализа речи осуществляется проекция в базис собственных векторов. Рассмотрим обратную задачу. Пусть имеется квазигармонический спектр (3), тогда сумма j-х квазигармоник данного спектра будет равна исходному фонемному ряду:

$$S_N = \sum_{n=0}^{L\text{-}1} \mathbf{T}_j^{<n>}, \; j = 0,\ldots,N-1. \qquad (4)$$

Пусть для некоторой последовательности i=0,1,… собственные числа λ_i, $\mathbf{u}^{<i>}$, $\mathbf{x}^{<i>}$ – эмпирически найденные величины, образуют совокупность параметров для образования звуков речи, тогда для произведения:

$$\mathbf{A}_i = \sqrt{\lambda_i}\,\mathbf{u}^{<i>}[\mathbf{x}^{<i>}]^{\mathrm{T}}, \; i=0,\ldots,$$

выражение (3) можно принять в качестве синтезатора акустических сигналов, генерируемых речеобразующим трактом (рисунок 1). Без решения прямой задачи синтезирование параметров λ_i, $\mathbf{u}^{<i>}$, $\mathbf{x}^{<i>}$ в качестве резонаторов речеобразующего тракта является достаточно сложным процессом [16, 17]. Тем не менее, систему:

$$\begin{cases}(3);\\(4).\end{cases} \tag{5}$$

можно принять в качестве сингулярной модели вокализированного сегмента речевого сигнала для решения задачи оценивания ЧОТ.

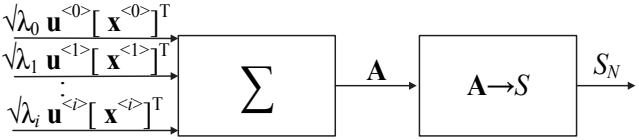

Рис. 1. Модель сингулярного синтезатора речи

Таким образом, можно сформулировать следующие фундаментальные тезисы для сингулярной модели вокализированной речи:

1. Система (5) наглядным образом показывает, что принимаемая сингулярная модель вокализированного сегмента речевого сигнала позволяет анализировать (рассматривать) речевой сигнал, в котором неизвестны амплитуды, периоды и начальные фазы всех гармоник.

2. Если речеобразующий тракт рассматривать как систему акустических резонаторов, тогда каждая *i*-я тройка чисел (λ_i, $\mathbf{u}^{<i>}$, $\mathbf{x}^{<i>}$), как отдельный параметр *i*-го резонатора, содержит информацию об индивидуальном акустическом различие, так как пространство собственных векторов **x** образует нестационарных базис, в который проецируется **A**.

3. При $i \rightarrow L$, модель (5) позволяет учитывать особенности речевосприятия через (3), а речеобразования через (4).

Теперь, исходя из (5), модель сингулярного оценивания ЧОТ можно представить в следующем концептуальном виде (рисунок 2):

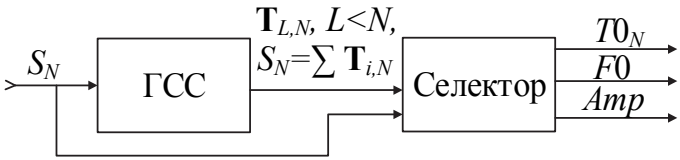

Рис. 2. Концептуальная модель сингулярного оценивания ЧОТ: S_N – входной сигнал; $\mathbf{T}_{L,N}$ – спектр временных рядов; ГСС – генератор сингулярного спектра; S_N – входной сигнал; $T0$ – трек основного тона; $F0$ – ЧОТ; Amp – амплитуда

1) средство генерации сингулярного спектра речевого сигнала, в котором входные данные – это фонемный ряд S_N, а выходные данные – это спектр временных рядов $\mathbf{T}_{L,N}$;

2) средство выбора спектральной составляющей соответствующей частоте основного тона речи, в котором входные данные – это спектр временных рядов $\mathbf{T}_{L,N}$, а выходные данные – это частота основного тона речи $F0$, средняя амплитуда Amp и квазигармоническая составляющая основного тона речи $T0$.

Численная реализация модели сингулярного оценивания ЧОТ речи выражается в системе:

$$
\begin{cases}
\mathbf{A} = [S_{i\text{-}1},\ldots,S_{i+L\text{-}1}]^{\mathrm{T}}, 1 \le i \le K, K = N - L + 1; \\
\mathbf{C} = \mathbf{A}\mathbf{A}^{\mathrm{T}}; \\
(\mathbf{U}_{\mathrm{C}}, \mathbf{D}_{\mathrm{C}}) = \mathrm{Eigens}(\mathbf{C}); \\
\mathbf{V}_{\mathrm{A}}{}^{\mathrm{T}} = \mathrm{D}_{\mathrm{C}}{}^{-1}\mathbf{U}_{\mathrm{C}}{}^{\mathrm{T}}\mathbf{A}; \\
(3); \\
f_n = \\
\dfrac{p}{N\Delta t}, p = \{k, \left\| [\dfrac{1}{N}\sum_{j=1}^{N} \mathbf{T}_j^{<n>} e^{\frac{-2\pi \mathbf{i}}{N}kj}]_k \right\| \subseteq MAX, k = \overline{1,N}\}, \\
n = \overline{1,L}; \\
f_j = f_n \in [f_{\min} \le f_n \le f_{\max}], n = \overline{1,L}, \\
j = 0,1,\ldots,K < L; \\
f_0 = f_{j=\text{нкчот}} = f_j \in \{\min(f_j), 2\min(f_j), \ldots, M\min(f_j)\}\}, \\
j = \overline{1,K}; \\
T0_n = T_{j=\text{нкчот},n}, n = \overline{1,N}; \\
F0 = \dfrac{1}{m-1}\sum_{i=1}^{m} \dfrac{1}{(k_{i-1} - k_i)\Delta t}, \\
k_i = \{n, T0_n \subset MAX, n = \overline{0,N-1}\}, i = \overline{1,m}; \\
Amp = \dfrac{1}{m}\sum \max(T0_n), n = 1,2,\ldots,m.
\end{cases}
$$

где: S_N – исходный временной ряд;
N – длина ряда;
L – размер спектрального окна;
$\mathbf{A}$ – траекторная (Ганкелева / H. Hankel matrix) матрица наблюдений [11];
$\mathbf{C}$ – бисимметричная матрица;
$\mathbf{U}_{\mathrm{C}}$ – левая сингулярная матрица поворота;

$\mathbf{V}_A{}^T$ – правая сингулярная матрица поворота;
$\mathbf{u}^{<n>}$ – левый сингулярный вектор;
$\mathbf{v}^{<n>}$ – правый сингулярный вектор ($\mathbf{v}^{<n>}=\mathbf{x}^{<n>}\in\mathbf{V}$);
D – диагональная матрица, состоящая из собственных значений λ_i бисимметричной матрицы **C**, при условии:

$$\mathbf{D}_C = \mathrm{diag}\{\lambda_0,\lambda_1,\lambda_2,...,\lambda_{L-1}\},\ \lambda_0 < \lambda_1 <,...,< \lambda_{L-1};$$

$\mathbf{T}_i^n$ – спектр временных рядов (квазигармонический спектр);
Eigens – функция поиска собственных чисел;
НКЧОТ – номер компоненты с частотой основного тона;
$\mathbf{T}_{j=\text{нкчот},N}$ – активация квазигармоники с НКЧОТ;
f_n – одномерное, частотное представление временного спектра $\mathbf{T}_i^n$ при условии, что $f_0\in[f_{\min}, f_{\max}]$, где f_0 – искомая частота основного тона такая, что:

$$f_0 \in \{\min(f_i),2\min(f_i),\ldots,M\min(f_i)\}$$

наименьшая кратная величина частоты;
p – индекс элемента в ряде T_i^n , соответствующий максимальной амплитуде от преобразований Фурье в n-й квазигармонике;
Δt – величина обратная частоте дискретизации;
$T0_N$ – временной ряд, соответствующий квазигармонике с частотой основного тона речи;
$F0$ – средняя частота основного тона речи такая, что:

$$F0 = \frac{f_0^1 + f_0^2 + ... + f_0^m}{m-1},$$

где (m-1) – число обратных величин равных периодам умещающихся в ряде $T0_N$ (f_0^i – локальная частота тона):

$$f_0^1 + f_0^2 + ... + f_0^m = \frac{1}{(k_2-k_1)\Delta t} + \frac{1}{(k_3-k_2)\Delta t} + \ldots$$

$$+\frac{1}{(k_m-k_{m-1})\Delta t} = \sum_{i=1}^{m}\frac{1}{(k_i-k_{i-1})\Delta t},$$

где k_i – номер индекса в точке максимума:

$$k_i = \{n, T0_n \subset max,\ n = \overline{0, N-1}\},\ i = \overline{1, m};$$

Amp – средняя амплитуда квазигармоники основного тона речи.

3. Описание программной реализации сингулярного оценивания частоты основного тона. На примере модели "черный ящик" [18–20] рассмотрим программный комплекс сингулярного оценивания ЧОТ, состоящий из 10-ти программно реализованных модулей (рисунок 3):

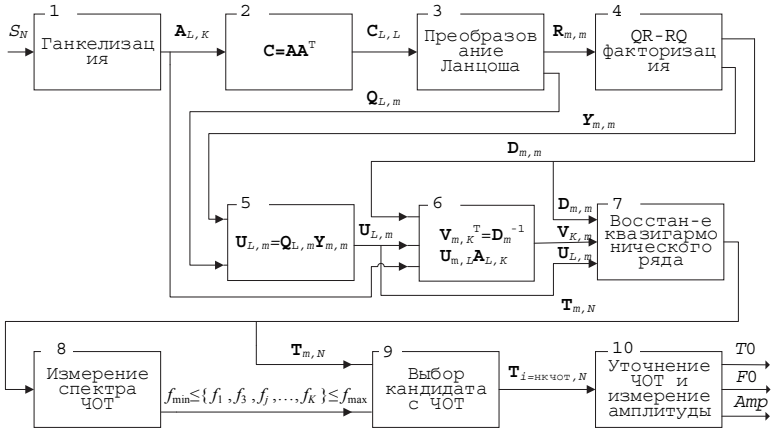

Рис. 3. Структура программного комплекса на уровне блоков

1) модуль Ганкелизации речевого сигнала S_N для получения траекторной матрицы $\mathbf{A}_{L,K}$;

2) модуль вычисления ковариационной матрицы $\mathbf{C}_{L,L}=\mathbf{A}\mathbf{A}^{\mathrm{T}}$;

3) модуль преобразований Ланцоша для вычисления трехдиагональной матрицы Релея $\mathbf{R}$ размерностью $m\times m$ и ортонормированного базиса подпространства Крылова $\mathbf{Q}_{L,m}$ [21, 22];

4) модуль QR факторизации для отыскания собственных пар ($\mathbf{y}^{<n>}\in\mathbf{Y}_{m,m}$, $\lambda_n\in\mathbf{D}$) матрицы Релея $\mathbf{R}_{m,m}$, где: $\mathbf{R}=\mathbf{Y}\,\mathbf{D}\,\mathbf{Y}^{\mathrm{T}}$, $\mathbf{Y}$–матрица собственных векторов матрицы $\mathbf{R}$, $\mathbf{D}$ – матрица собственных значений матрицы $\mathbf{R}$;

5) модуль вычисления первых m собственных векторов $\mathbf{u}^{<n>}\in\mathbf{U}$ (поиск матричной пары Ритца ($\mathbf{U}_{L,m}$, $\mathbf{D}_{m,m}$), где $\mathbf{U}_{L,m}=\mathbf{Q}_{L,m}\mathbf{Y}_{m,m}$ – матрица, состоящая из векторов Ритца);

6) модуль вычисления первых m собственных векторов $\mathbf{v}^{<n>}\in\mathbf{V}$ (матрицы главных компонент) траекторной матрицы $\mathbf{A}$, порождаемых ее строками:

$$\mathbf{V}_{m,K}^{\mathrm{T}} = \mathbf{D}_{m}^{-1}\mathbf{U}_{m,L}\mathbf{A}_{L,K};$$

7) модуль реконструкции первых m компонент квазигармонического спектра $\mathbf{T}_{m,N}$ речевого сигнала;

8) модуль измерения частоты квазигармонического спектра $\mathbf{T}_{m,N}$ (блок измерения частоты временного спектра);

9) модуль выбора кандидата с ЧОТ (блок выбора номера компоненты с частотой основного тона);

10) модуль уточнения ЧОТ и измерения амплитуды (блок вычисления частоты и амплитуды основного тона).

В соответствии с концептуальной моделью сингулярного измерителя, блоки 1-7 описывают генератор сингулярного спектра (ГСС), а блоки 8-10 описывают средство выбора квазигармонической составляющей основного тона и дальнейшего уточнения частоты и амплитуды основного тона (Селектор).

Таким образом, программная реализация сингулярного измерителя ЧОТ включает:

1. Конструктор класса генератора сингулярного спектра, реализованный в качестве функции (рисунок 4):

$$T=ssg(S, N, L, m),$$

где: S – массив данных, содержащий исходный фонемный ряд S_N;
N – размер массива S;
L – размер окна анализа;
m – число квазигармонических составляющих;
T – массив данных, содержащий квазигармонический спектр.

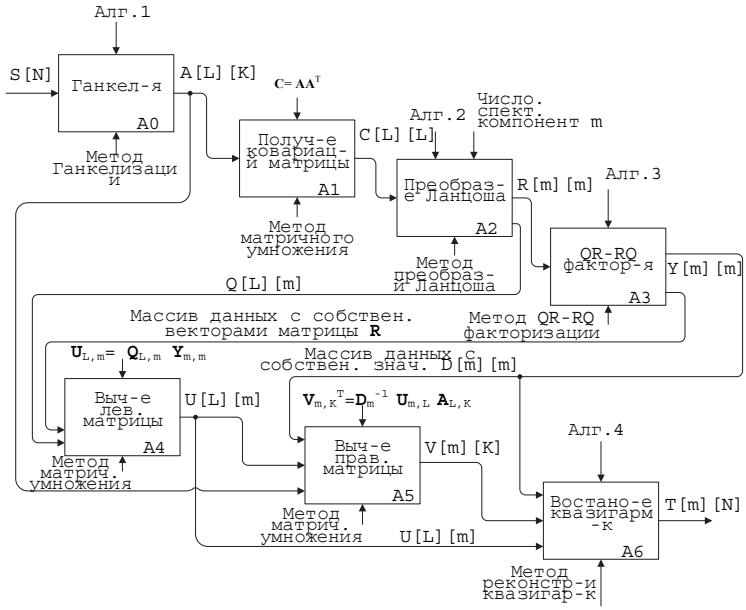

Рис. 4. Программная реализация генератор сингулярного спектра на уровне модели IDEF0

2. Модуль преобразований Ланцоша, реализованный в качестве метода класса ssg:

$$(R, Q)= \text{Lanczos}(C, RS, CS),$$

где: C – массив данных, содержащий ковариационную матрицу;
RS, CS – параметры, задающие размеры ковариационной матрицы **C**;
R – массив данных, содержащий трехдиагональную симметричную матрицу $\mathbf{R}_{m,m}$;
Q – массив данных, содержащий векторы Ланцоша матрицы $\mathbf{Q}_{L,m}$.

3. Модуль QR факторизации, реализованный в качестве метода класса ssg:

$$(D, Y)=\text{qr}(a, b, RS),$$

где:
a – массив данных, содержащий элементы трехдиагональной матрицы **R**, расположенных на главной диагонали;
b – массив данных, содержащий элементы трехдиагональной матрицы **R**, расположенных над главной диагональю;
RS –входной параметр, задающий размер массива a (количество спектральных компонент RS=*m*);
Y – массив данных, содержащий собственные векторы матрицы **R**.

4. Конструктор класса селектора, реализованный в качестве функции (рисунок 5):

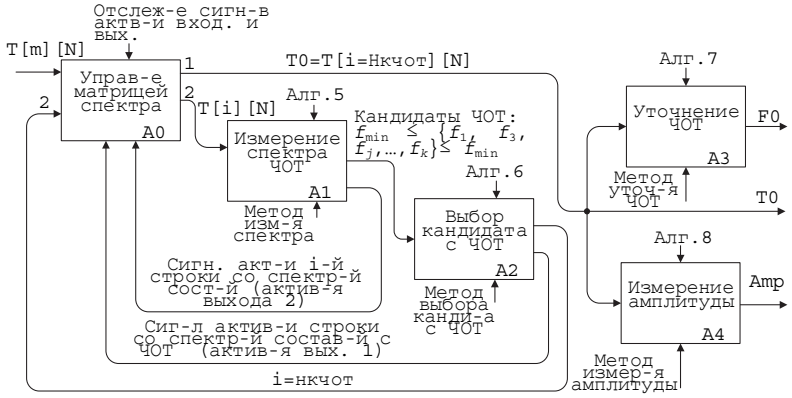

Рис. 5. Программная реализация селектора на уровне модели IDEF0

Selector(S, T, T0, F0, Amp, m, N),

где: T – массив данных, содержащий квазигармонический спектр;
T0 – массив данных, содержащий квазигармонику основного тона;
F0 – переменная, содержащая ЧОТ;
Amp – переменная, содержащая среднюю амплитуду квазигармоники основного тона.

4. Тестирование программной реализации сингулярного оценивания частоты основного тона. Рассмотрим общий вид работы программной реализации сингулярного оценивания ЧОТ. На вход программы подаются данные в виде фонемного ряда S, который выступает в качестве входного параметра для инициализации класса ГСС. Конструктор класса ГСС вызывает методы, в которых реализованы алгоритмы сингулярного спектрального анализа. В процессе работы вызываемых методов, осуществляется преобразование фонемного ряда, содержащегося в массиве данных S, в спектр квазигармоник, содержащихся в двухмерном массиве данных T. Массив данных T выступает в качестве входного аргумента при инициализации класса селектора. Конструктор класса селектора вызывает методы выбора квазигармонической составляющей с ЧОТ, и методы расчета его параметров. На выходе программы данные, соответствующие параметрам основного тона T0, Amp, F0.

Оценка временных характеристик сингулярного оценивания ЧОТ проводилась на персональном компьютере на базе процессора Intel i5 3.1GHz и мобильном устройстве связи на базе процессора Apple A6 1.7GHz (таблица 1). В качестве положительного критерия временных характеристик оценивания ЧОТ принималась работа программы в режиме реального времени. Под режимом реального времени понимается время сингулярного оценивания ЧОТ меньшее, чем сам кадр анализа. В качестве входных данных выбирались фонемные ряды гласных звуков русской речи, мужского и женского диктора, длительностью 32мс. В таблице 1 параметр G задает количество спектральных составляющих, которые необходимо найти, а ε задает достаточную ошибку округления для сингулярных чисел. Для уменьшения латентности анализа подбираются соответствующие параметры G и ε. Результаты тестирования временных характеристик показывают, что время оценивания ЧОТ (выполнения программы) как для ПК (Intel i5

3.1Ghz), так и для мобильного устройства связи (Apple A6 1.7Ghz) не превышает заданного начальным условием.

Таблица 1. Временные характеристики оценивания ЧОТ

Диктор № 2, пол: Ж, G=32									
Фонема	нк чот	Intel i5 3.1GHz				Apple A6 1.7GHz			
		Время (мс)	ЧОТ (Гц) при ε= 0,00001	Время (мс)	ЧОТ (Гц) при ε= 0,0001	Время (мс)	ЧОТ (Гц) при ε= 0,00001	Время (мс)	ЧОТ (Гц) при ε= 0,0001
[ɑ]	1	15	199,804	9	199,804	28	199,804	18	199,800
[e]	2	17	195,047	7	195,047	27	195,047	20	195,000
[ɵ]	2	15	199,804	9	199,804	29	199,804	25	199,800
[i]	1	16	204,800	8	204,800	27	204,800	22	204,800
[o]	2	17	210,051	10	210,051	26	210,051	21	204,800
[u]	1	18	215,578	6	215,578	29	215,578	20	215,570
[ɨ]	1	15	204,800	7	204,800	28	204,800	20	204,800
[ɪ]	2	18	199,804	9	199,804	30	199,804	18	199,800
[ʉ]	1	17	204,800	10	204,800	38	204,800	21	204,800
[æ]	1	16	186,181	7	186,181	27	186,181	19	186,180

Оценка точности сингулярного оценивания ЧОТ проводилась при следующих условиях:

1. Для известных алгоритмов RAPT, YIN, SWIPE', SHS, AC-P, AC-S, ANAL, CC, CEP, ESRPD, SHR, TEMPO [4-7, 23-29] и сингулярного оценивания ЧОТ (SEPT – Singular Estimation Pitch Tracking) рассматривался процент грубых ошибок GPE (gross pitch errors) [9]. Величина GPE показывает отношение количества анализируемых фреймов с отклонением полученной оценки ЧОТ более чем на ±20% от реального значения ЧОТ к общему числу вокализированных фреймов:

$$GPE(\%) = \frac{N_{\mathrm{GPE}}}{N_{\mathrm{V}}} 100,$$

где: N_{GPE} – число фреймов с отклонением полученной оценки более чем на ±20% от настоящего значения ЧОТ;
N_{V} – общее число вокализированных фреймов.

На первый взгляд 20%-я погрешность ошибки кажется слишком большой, но, учитывая, что большинство ошибок, допускаемых алгоритмами при оценивании ЧОТ варьируется в пределах октавы, то выбор такой погрешности можно считать обоснованным.

2. Доступ к известным алгоритмам осуществлялся с помощью программного обеспечения SFS [30], Praat [31], Straight [32], Aubio [33], Festival [34], SPE [35], (таблица 2).

3. В качестве анализируемого материала были выбраны речевые базы Disordered Voice Database (DVD) [36], Keele Pitch Database (KPD) [37] и Paul Bagshaw's Database (PBD) [38].

Если принять, что реализация всех алгоритмов выполнена в соответствии с их оригинальным описанием [4-7, 23-29], то при использовании идентичных входных данных (речевых фрагментов из выбранных баз) и единого аппаратного обеспечения (ПК на базе Intel i5 3.1GHz) можно считать, что сравнение алгоритмов проводились в идентичных условиях.

Таблица 2. Перечень алгоритмов оценивания ЧОТ и доступ к ним

Метод	Алгоритм	Программа (библиотека)	Доступ (функция)
Автокорреляционный	AC-S	SFS	fxac
Автокорреляционный	ANAL	SFS	fxanal
Кепстральный	CEP	SFS	fxcep
Кросскорреляционный	RAPT	SFS	fxrapt
Автокорреляционный	AC-P	PRAAT	ac
Кросскорреляционный	CC	PRAAT	cc
Спектральный	SHS	PRAAT	shs
Кросскорреляционный	ESRPD	FESTIVAL	pda
Спектральный	SHR	MATLABCENT.	shrp
Корреляционный	SWIPE'	SPE	Swipe
Спектральный	TEMPO	STRAIGHT	exstraightsource
Автокорреляционный	YIN	AUBIO	Aubiopitch

Таблица 3 показывает характеристику GPE для несортированных образцов вокализированных сегментов речи, выбранных из базы DVD (рисунок 6).

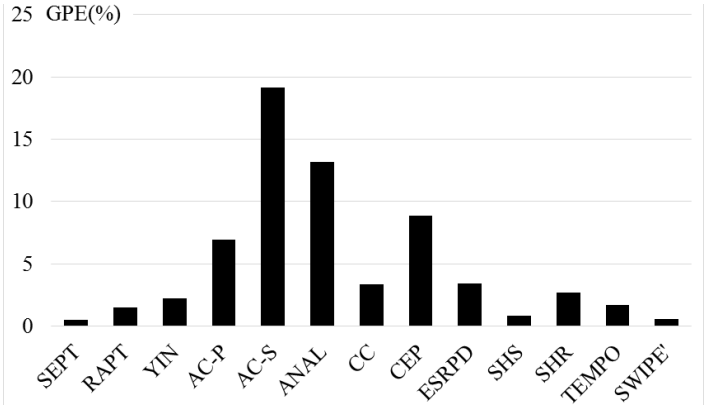

Рис. 6. Средний процент грубых ошибок оценки ЧОТ, допускаемый известными алгоритмами и SEPT

Таблица 3. Оценка грубых ошибок для натуральной речи

	Процент грубых ошибок - GPE (%)			
Алгоритм	База PBD	База KPD	База DVD	Среднее
SEPT	0,11	0,65	0,74	0,49
RAPT	0,78	1,08	2,70	1,52
YIN	0,35	1,43	4,90	2,22
AC-P	0,72	3,01	17,02	6,91
AC-S	8,90	7,40	41,18	19,16
ANAL	0,81	2,70	36,05	13,18
CC	0,45	3,65	6,03	3,37
CEP	6,20	4,23	16,07	8,83
ESRPD	1,35	3,99	5,00	3,44
SHS	0,16	1,03	1,34	0,84
SHR	0,71	1,56	5,80	2,69
TEMPO	0,33	1,97	2,91	1,73
SWIPE’	0,14	0,87	0,80	0,60

Таблица 4 показывает характеристику GPE в зависимости от пола диктора для баз PBD и KPD, т.к. для них имеются контрольные оценочные значения ЧОТ, полученные с помощью Ларинографа.

Таблица 4. Оценка грубых ошибок по полу

	Процент грубых ошибок - GPE (%)		
Алгоритм	Мужчины	Женщины	Среднее
SEPT	0,32	2,10	1,21
RAPT	0,45	2,81	1,63
YIN	1,19	3,12	2,16
AC-P	2,25	3,55	2,90
AC-S	3,17	9,40	6,29
ANAL	1,41	5,73	3,57
CC	2,54	4,43	3,49
CEP	2,00	4,17	3,09
ESRPD	3,20	3,79	3,50
SHS	0,62	2,39	1,51
SHR	0,68	3,57	2,13
TEMPO	0,75	2,98	1,87
SWIPE’	0,40	2,32	1,36
Среднее	1,97	3,81	2,89

На первый взгляд, величина GPE показывает степень робастности оценивания ЧОТ, так как, по сути, показывает процент допущенных ошибок каждым алгоритмов в процессе оценивания, но с другой стороны по данной величине можно судить о степени точности оценки

ЧОТ. Так, например, можно сказать, что для несортированной базы DVD, SEPT на 20% робастен по отношению к SWIPE', так как более точнее осуществляет оценку ЧОТ за счет сингулярной модели речевого сигнала, которая позволяет рассматривать речеобразующий тракт как систему акустических резонаторов, в которой параметрами выступают собственные значения и собственные векторы, содержащие информацию о структуре речевого сигнала с учетом нестационарных амплитуд, периодов и фаз гармоник, входящих в его состав.

5. Заключение. Время оценивания ЧОТ для 100 несортированных вокализованных образцов речи для ПК, при заданном ε=0,00001, не превышало 20мс. Таким образом, можно заключить, что сингулярное оценивание ЧОТ можно использовать в приложениях реального времени, где задержка в 20мс может быть допустимой. Результаты эксперимента показывают, что сингулярный метод оценивания ЧОТ более робастен по отношению к известным аналогам, а значит более точнее оценивает ЧОТ.

Литература

1. *Голубинский А.Н.* Оценка частоты основного тона речевого сигнала при априори неизвестных амплитудах и начальных фазах полигармонического несущего колебания // Вестник Воронежского института МВД России. 2010. № 3. С. 110–117.
2. *Ронжин А.Л., Басов О.О.* Определение степени алкогольной интоксикации человека на основе автоматического анализа речи // Вестник Московского университета МВД России. 2015. № 5. С. 216–220.
3. *Meshcheryakov R.V., Balatskaya L.N., Choinzonov E.L., Chizevskaya S.Yu., Kostyuchenko E.U.* Software for Assessing Voice Quality in Rehabilitation of Patients after Surgical Treatment of Cancer of Oral Cavity, Oropharynx and Upper Jaw // Proceedings of 15th International Conference SPECOM 2013. Pilsen. Czech Republic. 2013. pp 294–301.
4. *Talkin D.* A Robust Algorithm for Pitch Tracking (RAPT) // Speech Coding & Synthesis. 1995. pp-495–518.
5. *Cheveigne A., Kawahara H.* YIN, a fundamental frequency estimator for speech and music // Jour. Acoust. Soc. Am. 2002. vol. 111. no. 4. pp. 1917–1930.
6. *Camacho A., Harris J.G.* A sawtooth waveform inspired pitch estimator for speech and music // Journal Acoust. Soc. Am. 2008. vol. 123. no. 4. pp. 1638–1652.
7. *Hermes D.J.* Measurement of pitch by subharmonic summation // Jour. Acoust. Soc. Am. 1988. vol. 83. pp. 257–264.
8. *Rabiner L.R., Schafer R.W.* Digital processing of speech signals // Prentice Hall. 1978.
9. *Azarov E., Vashkevich M., Petrovsky A.* Instantaneous pitch estimation based on RAPT framework // Proceedings of the 20th European Signal Processing Conference (EUSIPCO). Bucharest. 2012. pp. 2787–2791.
10. *Basov O.O., Ronzhin A.L., Budkov V.Yu.* Optimization of Pitch Tracking and Quantization // Proc. SPECOM-2015. LNAI 9319. 2015. pp. 65–72.
11. *Basov O.O., Ronzhin A.L., Budkov V.Yu.,* Saitov *I.A.* Method of Defining Multimodal Information Falsity for Smart Telecommunication Systems // Internet of Things, Smart Spaces, and Next Generation Networks and Systems. Springer. St. Petersburg. Russia. 2015. LNCS 9247. pp. 163–173.

12. *Golyandina N., Zhigljavsky A.* Singular Spectrum Analysis for time series // Springer Science & Business Media. 2013.
13. *Tony F.C.* An Improved Algorithm for Computing the Singular Value Decomposition // ACM Transaction on Mathematical Software. 1982. vol. 8. no. 1. pp. 72–83.
14. *Азаров И.С., Вашкевич М.И., Лихачев Д.С., Петровский А.А.* Изменение частоты основного тона речевого сигнала на основе гармонической модели с нестационарными параметрами // Труды СПИИРАН. 2014. Вып. 32. С.5–26.
15. *Бондаренко В.П., Коцубинский В.П., Мещеряков Р.В.* Нестационарные модели в обработке речевых сигналов // Акустика речи. Медицинская и биологическая акустика. Архитектурная и строительная акустика и вибрации. Сб. трудов XVIII сессии Российского акустического общества. М.: ГЕОС. 2006. Т.3. С. 8–11.
16. *Вольф Д.А.* Спектральная теорема для решения частичной проблемы собственных чисел степенным методом в задачах сингулярного спектрального анализа речи // Системы управления и информационные технологии. 2014. №3.1(57). С. 129–135.
17. *Агаев Р.П., Чеботарев П.Ю.* Метод проекции в задаче о консенсусе и регуляризованный предел степеней стохастической матрицы //Автоматика и телемеханика. 2011. №. 12. С. 38–59.
18. *Налимов В.В.* Теория эксперимента // М.: Наука. 1971. 208 с.
19. *Силич В.А., Комагоров В.П., Савельев А.О.* Принципы разработки системы мониторинга и адаптивного управления разработкой «интеллектуального» месторождения на основе постоянно действующей геологотехнологической модели // Известия Томского политехнического университета. 2013. Т. 323. №. 5. С. 94–100.
20. *Силич В. А., Силич М.П., Аксенов С.В.* Алгоритм построения нечеткой системы логического вывода Мамдани, основанный на анализе плотности обучающих примеров //Доклады томского государственного университета систем управления и радиоэлектроники. 2013. №. 3(29). С. 76–82.
21. *Parlett B. N.* The symmetric eigenvalue problem // Englewood Cliffs. NJ: Prentice-Hall. 1980. vol. 7.
22. *Knizhnerman L., Simoncini V.* A new investigation of the extended Krylov subspace method for matrix function evaluations // Numerical Linear Algebra with Applic. 2010. vol. 17. no. 4. pp. 615–638.
23. *Boersma P.* Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound // Proceedings of the institute of phonetic sciences. 1993. vol. 17. no. 1193. pp. 97–110.
24. *Secrest B. G., Doddington G. R.* An integrated pitch tracking algorithm for speech systems // Acoustics, Speech, and Signal Processing. IEEE International Conference on ICASSP'83. 1983. vol. 8. pp. 1352–1355.
25. *Noll A. M.* Cepstrum pitch determination // The journal of the acoustical society of America. 1967. vol. 41. no. 2. pp. 293–309.
26. *Bagshaw P. C., Hiller S. M., Jack M. A.* Enhanced pitch tracking and the processing of F0 contours for computer and intonation teaching // Proc. Europe-an Conf. on Speech Comm. (Eurospeech). 1993. pp. 1003–1006.
27. *Medan Y., Yair E., Chazan D.* Super resolution pitch determination of speech signals // IEEE Trans. Signal Process. 1991. vol. 39. pp. 40-48.
28. *Sun X.* A pitch determination algorithm based on subharmonic-to-harmonic ratio // The 6th International Conference of Spoken Language Processing. 2000. pp. 676–679.
29. *Kawahara H., Katayose H., de Cheveigne A., Patterson R. D.* Fixed Point Analysis of Frequency to Instantaneous Frequency Mapping for Accurate Esti-mation of F0 and Periodicity // Proc. EUROSPEECH. 1999. vol. 99. Issue 6. pp. 2781–2784.

30. Speech Filing System (SFS) // UCL Psychology & Language sciences Faculty of Brain Sciences. 2015. URL: http://www.phon.ucl.ac.uk/resource/sfs/ (дата обращения: 17.09.2015).
31. Praat // Phonetic Sciences. Amsterdam. 2015. URL: http://www.fon.hum.uva.nl/praat/download_win.html (дата обращения: 17.09.2015).
32. Straight // GitHub. 2015. URL: https://github.com/shuaijiang/STRAIGHT (дата обращения: 17.09.2015).
33. Aubio //Aubio. 2015. URL: http://aubio.org/download (дата обращения: 17.09.2015).
34. Festival // The Festival Speech Synthesis System. 2015. URL: http://www.cstr.ed.ac.uk/projects/festival/download.html (дата обращения: 17.09.2015).
35. SWIPE' pitch estimator. 2015. URL: https://github.com/kylebgorman/swipe (дата обращения: 17.09.2015).
36. Disordered Voice Database. 2015. URL: http:// http://kayelemetrics.com (дата обращения: 17.09.2015).
37. Keele Pitch Database. 2015. URL: http://www.icocla.it/keele.html (дата обращения: 20.03.2015).
38. Paul Bagshaw's Database. 2015. URL: http://www.cstr.ed.ac.uk/research/projects/fda (дата обращения: 17.09.2015).

References

1. Golubinsky A.N. [Pitch frequency estimation of a speech signal at a priori unknown amplitudes and initial phases of polyharmonical carrying oscillation]. *Vestnik Voronezhskogo instituta MVD Rossii – Vestnik of Voronezh Institute of the Ministry of the Interior of Russia*. 2010. vol.3. pp. 110–117. (In Russ.).
2. Ronzhin A.L., Basov O.O. [Detection of alcohol intoxication degree based on automatic speech analysis]. *Vestnik Moskovskogo universiteta MVD Rossii – Herald of Moscow University of the MIA of Russia*. 2015. no. 5. pp. 216–220. (In Russ.).
3. Meshcheryakov R.V., Balatskaya L.N., Choinzonov E.L., Chizevskaya S.Yu., Kostyuchenko E.U. Software for Assessing Voice Quality in Rehabilitation of Patients after Surgical Treatment of Cancer of Oral Cavity, Oropharynx and Upper Jaw. Proceedings of 15th International Conference SPECOM 2013. Pilsen. Czech Republic. 2013. pp 294–301.
4. Talkin D. A Robust Algorithm for Pitch Tracking (RAPT). *Speech Coding & Synthesis*. 1995. pp-495–518.
5. Cheveigne A., Kawahara H. YIN, a fundamental frequency estimator for speech and music. *Jour. Acoust. Soc. Am*. 2002. vol. 111. no. 4. pp. 1917–1930.
6. Camacho A., Harris J.G. A sawtooth waveform inspired pitch estimator for speech and music. *Journal Acoust. Soc. Am*. 2008. vol. 123. no. 4. pp. 1638–1652.
7. Hermes D.J. Measurement of pitch by subharmonic summation. *Jour. Acoust. Soc. Am*. 1988. vol. 83. pp. 257–264.
8. Rabiner L.R., Schafer R.W. Digital processing of speech signals. Prentice Hall. 1978.
9. Azarov E., Vashkevich M., Petrovsky A. Instantaneous pitch estimation based on RAPT framework. Proceedings of the 20th European Signal Processing Conference (EUSIPCO). Bucharest. 2012. pp. 2787–2791.
10. Basov O.O., Ronzhin A.L., Budkov V.Yu. Optimization of Pitch Tracking and Quantization. Proc. SPECOM-2015. LNAI 9319. 2015. pp. 65–72.
11. Basov O.O., Ronzhin A.L., Budkov V.Yu., Saitov I.A. Method of Defining Multimodal Information Falsity for Smart Telecommunication Systems. Internet of Things, Smart Spaces, and Next Generation Networks and Systems. Springer. St. Petersburg. Russia. 2015. LNCS 9247. pp. 163–173.

12. Golyandina N., Zhigljavsky A. Singular Spectrum Analysis for time series. Springer Science & Business Media. 2013.
13. Tony F.C. An Improved Algorithm for Computing the Singular Value Decomposition. ACM Transaction on Mathematical Software. 1982. vol. 8. no. 1. pp. 72–83.
14. Azarov E., Vashkevich M., Likhachov D., Petrovsky A. Pitch modification of speech signal using harmonic model with time-varying parameters. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2014. vol. 32. pp.5–26. (In Russ.).
15. Bondarenko V.P., Kotsubinsky V.P., Meshcheryakov R.V. [Non-stationary models in speech signal processing]. *Acistica rechi. Medicinskaja i biologicheskaja akustika. Arhitekturnaja i stroitel'naja akustika i vibracii. Sbornik trudov XVIII sessii Russain acoustic society* [Acoustics of speech. Medical and biological acoustics. Architectural and building acoustics and vibration. Proceedings of the 17th Session of the Russian Acoustical Society. M: GEOS. 2006. vol. 3. pp. 8–11. (In Russ.).
16. Volf D.A. [The use of spectral theorem for power-iteration solution of a partial eigenvalue problem in singular spetrum analysis of speech]. *Sistemy Upravleniia i Informatsionnye Tekhnologii – Control Systems and Information Technology*. 2014. no. 3.1(57). pp. 129–135. (In Russ.).
17. Agaev R.P., Chebotarev P.Yu. [The projection method for reaching consensus and the regularized power limit of a stochastic matrix]. *Avtomatika i Telemekhanika – Automation and Remote control*. 2011. vol. 12. pp. 38–59. (In Russ.).
18. Nalimov V.V. *Teorija jeksperimenta* [The theory of experiment]. 1971. 208 p. (In Russ.).
19. Silich V. A., Komagorov V. P., Savelev A. O. [Principles of developing the system of monitoring and adaptive controlling the intelligent oil field study based on permanent geological and technological models]. *Izvestija Tomskogo politehnicheskogo universiteta – Tomsk Polytechnic University*. 2013. vol. 323. no. 5. pp. 94–100. (In Russ.).
20. Silich V. A., Yampolsky V.Z., Savelyev A.O., Komagorov V.P., Alekseev A.A., Grebenshchikov S.A. [A Mamdani-type fuzzy system construction algorithm based on training vectors density analysis]. *Doklady tomskogo gosudarstvennogo universiteta sistem upravlenija i radiojelektroniki – Proceedings of Tomsk State University of Control Systems and Radioelectronics*. 2013. vol. 3(29). pp. 76-82. (In Russ.).
21. Parlett B. N. The symmetric eigenvalue problem. Englewood Cliffs. NJ: Prentice-Hall. 1980. vol. 7.
22. Knizhnerman L., Simoncini V. A new investigation of the extended Krylov subspace method for matrix function evaluations. *Numerical Linear Algebra with Applic*. 2010. vol. 17. no. 4. pp. 615–638.
23. Boersma P. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Proceedings of the institute of phonetic sciences. 1993. vol. 17. no. 1193. pp. 97–110.
24. Secrest B.G., Doddington G.R. An integrated pitch tracking algorithm for speech systems. Acoustics, Speech, and Signal Processing. IEEE International Conference on ICASSP'83. 1983. vol. 8. pp. 1352–1355.
25. Noll A.M. Cepstrum pitch determination. *The journal of the acoustical society of America*. 1967. vol. 41. no. 2. pp. 293–309.
26. Bagshaw P.C., Hiller S.M., Jack M.A. Enhanced pitch tracking and the processing of F0 contours for computer and intonation teaching. Proc. Europe-an Conf. on Speech Comm. (Eurospeech). 1993. pp. 1003–1006.
27. Medan Y., Yair E., Chazan D. Super resolution pitch determination of speech signals. *IEEE Trans. Signal Process*. 1991. vol. 39. pp. 40-48.
28. Sun X. A pitch determination algorithm based on subharmonic-to-harmonic ratio. The 6th International Conference of Spoken Language Processing. 2000. pp. 676–679.

29. Kawahara H., Katayose H., de Cheveigne A., Patterson R.D. Fixed Point Analysis of Frequency to Instantaneous Frequency Mapping for Accurate Esti-mation of F0 and Periodicity. Proc. EUROSPEECH. 1999. vol. 99. Issue 6. pp. 2781–2784.
30. Speech Filing System (SFS). UCL Psychology & Language sciences Faculty of Brain Sciences. Available at: http://www.phon.ucl.ac.uk/resource/sfs/ (accessed 17.09.2015).
31. Praat. Phonetic Sciences, Amsterdam. 2015. Available at: http://www.fon.hum.uva.nl/praat/download_win.html (дата обращения: 17.09.2015).
32. Straight. GitHub. Available at: https://github.com/shuaijiang/STRAIGHT (accessed 17.09.2015).
33. Aubio. Aubio. Available at: http://aubio.org/download (accessed 17.09.2015).
34. The Festival Speech Synthesis System. Available at: http://www.cstr.ed.ac.uk/projects/festival/download.html (accessed 17.09.2015).
35. SPE. SWIPE' pitch estimator. Available at: https://github.com/kylebgorman/swipe (accessed 17.09.2015).
36. Disordered Voice Database. Available at: http:// http://kayelemetrics.com (accessed 17.09.2015).
37. Keele Pitch Database. Available at: http://www.icocla.it/keele.html (accessed 20.03.2015).
38. Paul Bagshaw's Database. Available at: http://www.cstr.ed.ac.uk/research/projects/fda (accessed 17.09.2015).

Вольф Данияр Александрович — аспирант кафедры комплексной информационной безопасности электронно-вычислительных систем, Томский государственный университет систем управления и радиоэлектроники (ТУСУР). Область научных интересов: системный анализ, сингулярный анализ, программирование, моделирование. Число научных публикаций — 16. runsolar@mail.ru; пр. Ленина, 40, Томск, 634050; р.т.: +7(3822)900-111, Факс: +7(3822)900-111.

Volf Daniiar Aleksandrovich — Ph.D. student of complex security of electronic-computing systems department, Tomsk State University of Control Systems and Radioelectronics (TUSUR). Research interests: speech analysis, speech recognition, signal analisys. The number of publications — 16. runsolar@mail.ru; 40, Lenin-avenue Tomsk, 634050, Russia; office phone: +7(3822)900-111, Fax: +7(3822)900-111.

Мещеряков Роман Валерьевич — д-р техн. наук, доцент, профессор кафедры комплексной информационной безопасности электронно-вычислительных систем, Томский государственный университет систем управления и радиоэлектроники (ТУСУР). Область научных интересов: системный анализ, информационная безопасность, вопросы обработки информации в интеллектуальных системах, особое внимание уделяется вопросам создания информационно-безопасных систем. Число научных публикаций — 247. mrv@ieee.org; пр. Ленина, 40, Томск, 634050; р.т.: +7(3822)900111, Факс: +7 (3822) 900-111.

Meshcheryakov Roman Valerievich — Ph.D., Dr. Sci., professor, professor of complex security of electronic-computing systems department, Tomsk State University of Control Systems and Radioelectronics (TUSUR). Research interests: speech analysis, speech recognition, medical technology, information security. The number of publications — 247. mrv@ieee.org; 40, Lenin-avenue Tomsk, 634050, Russia; office phone: +7(3822)900111, Fax: +7(3822)900111.

Поддержка исследований. Работа выполнена в рамках государственного задания Томского государственного университета систем управления и радиоэлектроники (проект № 3657).

Acknowledgements. Tomsk State Unuversity of Control Systems and Radioelectronics (project 3657).

РЕФЕРАТ

Вольф Д.А. Мещеряков Р.В. **Модель и программная реализация сингулярного оценивания частоты основного тона речевого сигнала.**

В статье рассматривается сингулярная модель речевого сигнала, которая позволяет рассматривать речеобразующий тракт как систему акустических резонаторов, в которой параметрами выступают собственные значения и собственные векторы, содержащие информацию о структуре речевого сигнала с учетом нестационарных амплитуд, периодов и фаз гармоник, входящих в его состав. Данное свойство обусловлено тем, что пространство собственных векторов образует нестационарный базис, в который проецируется речевой сигнал. Однако повышение точности вычисления ЧОТ приводит к увеличению вычислительной сложности. Предлагаемая в статье модель сингулярного оценивания частоты основного тона позволяет оптимизировать временную обработку речевого сигнала за счет аппроксимации края сингулярного спектра, выделяя главные компоненты, образующие речевой сигнал для случая неизвестных априорных распределений амплитуд, периодов и начальных фаз гармоник. Далее рассматривается программная реализация модели и проводится сравнение с аналогами. В результате, в данной работе предлагается новый подход к оцениванию частоты основного тона речевого сигнала.

SUMMARY

Volf D.A., Meshcheryakov R. V. **Software Implementation of a Singular Meter of the Pitch Frequency of a Speech Signal.**

The article discusses a singular model of the speech signal, which allows considering speech production as a system of acoustic resonators in which the parameters are the eigenvalues and eigenvectors containing the information about the structure of the speech signal, taking into account time-dependent amplitudes, periods and phases of the harmonics included in its composition. This property is determined by the fact that the space of eigenvectors forms a transient basis, in which the speech signal is projected. Improving the accuracy of calculating the pitch frequency leads to higher computational complexity. The proposed in the article singular model of estimating the pitch frequency allows optimizing the time processing of the natural speech signal by approximating the edges of the singular spectrum, highlighting the main components that form a voice signal for the case of unknown a priori distributions of amplitudes, periods and the initial phases of the harmonics. In addition, software implementation of the model and a comparison with analogues are considered. As a result, this paper proposes a new approach to estimating the pitch frequency of the speech signal.