

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ИЗМЕРЕНИЯХ

УДК 004.934

DOI 10.21685/2307-5538-2018-4-9

А. К. Алимуратов, Ю. С. Квитка, П. П. Чураков, А. Ю. Тычков

ПОВЫШЕНИЕ ТОЧНОСТИ ИЗМЕРЕНИЯ ЧАСТОТЫ ОСНОВНОГО ТОНА НА ОСНОВЕ ОПТИМИЗАЦИИ ПРОЦЕССА ДЕКОМПОЗИЦИИ РЕЧЕВЫХ СИГНАЛОВ НА ЭМПИРИЧЕСКИЕ МОДЫ

A. K. Alimuradov, Yu. S. Kvitka, P. P. Churakov, A. Yu. Tychkov

INCREASING THE ACCURACY OF MEASURING THE PITCH FREQUENCY BASED ON THE OPTIMIZATION OF THE PROCESS OF DECOMPOSITION OF SPEECH SIGNALS ON EMPIRICAL MODES

А н н о т а ц и я. Актуальность и цели. Рассматривается актуальный вопрос – повышение точности измерения частоты основного тона речевых сигналов. Объектом исследования является декомпозиция речевых сигналов на детерминистические и стохастические частотные составляющие с применением различных методов. Предметом исследования является улучшенная полная множественная декомпозиция на эмпирические моды с адаптивным шумом. Целью исследования является оптимизация процесса декомпозиции для повышения точности измерения частоты основного тона за счет улучшения частотно-избирательных свойств. **Материалы и методы.** Для декомпозиции речевых сигналов на частотные составляющие применялась адаптивная технология анализа нестационарных сигналов – улучшенная полная множественная декомпозиция на эмпирические моды с адаптивным шумом. Исследования и анализ данных были выполнены в среде математического моделирования © Matlab (MathWorks). **Результаты.** Кратко представлен обзор известных методов декомпозиции, основанных на преобразовании Фурье и вейвлет-преобразовании. Рассмотрены их преимущества и недостатки, выявлены перспективы применения метода декомпозиции на эмпирические моды. Представлено подробное математическое описание разновидностей декомпозиций и подчеркнута необходимость оптимизации улучшенной полной множественной декомпозиции на эмпирические моды с адаптивным шумом. Представлен способ оптимизации улучшенной декомпозиции для повышения точности измерения частоты основного тона с кратким математическим описанием. **Выводы.** Проведены исследования способа, направленные для оптимального определения параметров функционирования улучшенной декомпозиции. Показано, что предлагаемый способ на основе улучшенной полной множественной декомпозиции на эмпирические моды с адаптивным шумом успешно решает задачу повышения точности измерения частоты основного тона за счет лучших частотно-избирательных свойств.

A b s t r a c t. Background. The article considers an urgent problem – increasing the accuracy of measuring the pitch frequency of speech signals. The object of research is the decomposition of speech signals on deterministic and stochastic frequency components using various

methods. The subject of research is an improved complete ensemble empirical mode decomposition with adaptive noise. The purpose of research is the optimization of process of decomposition to improve the accuracy of measuring the pitch frequency due to improvement of the frequency-selective properties. **Materials and methods.** For the decomposition of speech signals on frequency components, an adaptive technology for analyzing non-stationary signals -an improved complete ensemble empirical mode decomposition with adaptive noise was used. Research and analysis of the data was performed in the environment of mathematical modeling © Matlab (MathWorks). **Results.** A brief overview of known methods of decomposition based on the Fourier transform and wavelet transform is presented. Their advantages and disadvantages are considered, prospects of applying the empirical mode decomposition method are revealed. A detailed mathematical description of the varieties of decompositions is presented and the need of optimization of improved complete ensemble empirical mode decomposition with adaptive noise is emphasized. A method of optimization of improved decomposition for increasing the accuracy of measuring the pitch frequency with a brief mathematical description is presented. **Conclusions.** The research of the method aimed at optimal determination of the operating parameters for the improved decomposition is carried out. It is shown that the proposed method, based on the improved complete ensemble empirical mode decomposition with adaptive noise, successfully solves the problem of increasing the accuracy of measuring the pitch frequency due to the best frequency-selective properties.

К л ю ч е в ы е с л о в а: обработка речевых сигналов, полная множественная декомпозиция на эмпирические моды с адаптивным шумом, частота основного тона, оптимизация.

Key words: processing of speech signals, complete ensemble empirical mode decomposition with adaptive noise, pitch frequency, optimization.

Введение

Речь состоит из вокализованных и невокализованных участков, образующихся в результате периодических и непериодических колебаний голосовых связок. Периодические колебания голосовых связок, частота которых является важным информативным параметром речи, называется частотой основного тона (ЧОТ) [1].

Измерение ЧОТ является одной из наиболее важных задач в обработке речевых сигналов, точность которой определяется временным и частотным разрешением анализа. Выбор определенного способа измерения ЧОТ представляет собой некоторый компромисс между вычислительной сложностью и устойчивостью к ошибкам. Функционально все способы измерения ЧОТ включают в себя три этапа: предварительную обработку, выделение периода и измерение значения ЧОТ [2]. Для повышения точности измерения ЧОТ наибольший интерес представляет этап предварительной обработки, в рамках которого речевой сигнал раскладывается на детерминистические и стохастические частотные составляющие с применением различных методов декомпозиции [1, 2].

Наиболее популярными методами декомпозиции являются методы, основанные на преобразовании Фурье (ПФ) и вейвлет-преобразовании (ВП), которые обладают рядом преимуществ и недостатков [3].

Несмотря на то, что ПФ обладает относительной простотой вычислений и выполняется при довольно общих условиях, существует несколько ограничений, накладываемых на исследуемые речевые сигналы [1, 4]. Сигналы должны быть строго периодическими функциями или содержать так называемые повторяющиеся, не обязательно являющиеся идентичными, фрагменты одной общей природы, так как в противном случае анализ в частотной области является малоэффективным. Для повышения эффективности анализа нестационарных речевых сигналов необходим подход, обладающий свойством адаптивности к каждому конкретному сигналу. Требованию адаптивности удовлетворяет дискретное ВП [1, 5]. В силу особенностей материнского вейвлета появляется возможность адаптивно обрабатывать сигналы путем довольно точного учета локальных временных особенностей. Но главная проблема – сложность выбора материнского вейвлета для решения конкретной задачи из их большого многообразия – полностью не решена [6–9]. Можно использовать автоматический подбор вейвлет-функций на основе априорной информации об анализируемом сигнале, однако эта процедура оказыва-

ется затратной по вычислительным ресурсам, что делает невозможной обработку нестационарных речевых сигналов сложной формы [10].

Исследования методов обработки сигналов [11] выявили перспективность использования адаптивной технологии декомпозиции нестационарных данных – декомпозиции на эмпирические моды (ДЭМ) [12].

Материалы и методы

Декомпозиция на эмпирические моды и ее разновидности

ДЭМ – это адаптивный метод анализа нестационарных сигналов, возникающих в нелинейных системах, который обеспечивает локальное разложение сигнала на быстрые и медленные колебательные функции. В результате разложения исходный сигнал представляется в виде суммы амплитудно- и частотно-модулированных функций – эмпирических мод (ЭМ). Аналитическое выражение ДЭМ выглядит следующим образом:

$$x(n) = \sum_{i=1}^I IMF_i(n) + r_i(n),$$

где $x(n)$ – исходный сигнал; $IMF_i(n)$ – ЭМ; $r_i(n)$ – конечный остаток; $i = 1, 2, \dots, I$ – номер ЭМ; n – дискретный отсчет времени ($0 < n \leq N$), N – количество дискретных отсчетов в сигнале).

В результате разложения речевого сигнала с помощью ДЭМ в одной ЭМ могут оказаться несоизмеримые по амплитудному и частотному масштабам колебательные функции или, наоборот, соизмеримые колебательные функции могут оказаться в разных модах. Это явление называется смешиванием ЭМ вследствие перекрытия масштабно-энергетических пространств мод. Для его снижения предложена множественная ДЭМ (МДЭМ) [13], состоящая в добавлении к исходному сигналу белого шума для создания новых экстремумов функции

$$x_j(n) = x(n) + w_j(n),$$

где $x_j(n)$ – шумовые копии исходного сигнала; $w_j(n)$ – реализации белого шума с нулевой средней единичной дисперсией.

Аналитическое выражение МДЭМ выглядит следующим образом:

$$x_j(n) = \sum_{i=1}^I IMF_{ji}(n) + r_{ji}(n),$$

где $j = 1, 2, \dots, J$ – количество реализаций белого шума.

В результате разложения получаются более регулярные ЭМ с соизмеримыми масштабами колебательных функций. Основной недостаток МДЭМ – декомпозиция не является полной, и каждая шумовая копия исходного сигнала $x_j(n)$ разлагается независимо от других реализаций, а каждый остаток $r_{ji}(n) = r_{j,i-1}(n) - IMF_{ji}(n)$ вычисляется без связи между различными реализациями. Кроме этого, в ЭМ наблюдается остаточный белый шум, а различные реализации шумовых копий порождают разное количество ЭМ, что затрудняет окончательное усреднение.

Другая разновидность ДЭМ – комплементарная МДЭМ [14], которая решает проблему остаточного шума, добавляя и вычитая пары шумов с прямыми и инверсными значениями амплитуды. Однако проблема усреднения остается нерешенной, поскольку разные шумовые копии исходного сигнала также могут производить разное количество ЭМ:

$$\begin{bmatrix} x_j(n) \\ x_j^*(n) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \times \begin{bmatrix} x(n) \\ w_j(n) \end{bmatrix},$$

где $x_j(n)$ – шумовые копии исходного сигнала; $x_j^*(n)$ – шумовые копии исходного сигнала с инверсными значениями амплитуды шума.

Учитывая эти недостатки, в работе [15] предлагается новый метод, называемый полной МДЭМ с адаптивным шумом (ПМДЭМАШ). Основная идея метода заключается в добавлении к исходному сигналу контролируемого шума для создания новых экстремумов. Первая ЭМ извлекается по методике МДЭМ, усредняя первые моды сигнала с белым шумом:

$$\overline{IMF}_1(n) = \frac{1}{J} \sum_{j=1}^J IMF_{j1}(n) = \widetilde{IMF}_1(n).$$

Затем вычисляется первый остаток, независимый от реализации шума:

$$r_1(n) = x(n) - \widetilde{IMF}_1(n),$$

Для дальнейшего извлечения мод к текущему первому остатку добавляется конкретный шум. Этот шум представляет собой ЭМ белого шума, полученную по методике ДЭМ.

Несмотря на вышеупомянутые достоинства метода ПМДЭМАШ, в работе [14] авторы отмечают следующие недостатки данного метода: ЭМ содержат остаточный шум; информативные моды о сигнале при разложении извлекаются «позже», чем при МДЭМ с некоторыми «паразитными» модами на ранних этапах декомпозиции.

В работе [16] авторы решают и предлагают улучшенную ПМДЭМАШ. Суть уменьшения остаточного шума заключается в использовании локальных средних значений вместо значений ЭМ. Метод МДЭМ независимо разлагает каждую реализацию сигнала с шумом, поэтому на первом этапе каждой реализации декомпозиции имеется одно локальное среднее значение и одна мода. Учитывая, что истинная мода может быть определена как разность между текущим остатком и усредненным значением его локальных средних величин, получим следующее выражение:

$$E_1(x(n)) = x(n) - M(x(n)),$$

где $E(\cdot)$ – аппарат извлечения ЭМ методом ДЭМ; $M(\cdot)$ – аппарат, создающий локальное среднее значение применяемого сигнала.

Тогда для первых ЭМ, полученных методами МДЭМ и ПМДЭМАШ, выражение будет иметь следующий вид:

$$\widetilde{IMF}_1(n) = E_1(x_j(n)) = x_j(n) - M(x_j(n)) = x_j(n) - \langle x_j(n) \rangle,$$

где $\langle \rangle$ – действие усреднения.

Оценивая только локальное среднее значение и вычитая его из исходного сигнала, получим следующее выражение:

$$\widetilde{IMF}_1(n) = x(n) - M(x_j(n)).$$

Суть устранения возникающих паразитных ЭМ на ранних этапах декомпозиции заключается в уменьшении перекрытия масштабно-энергетических пространств первых двух мод.

Алгоритм и математическое описание улучшенной ПМДЭМАШ:

Шаг 1. С помощью аппарата ДЭМ, выражая из формулы $E_1(x_j(n)) = x_j(n) - M(x_j(n))$ локальные средние значения шумовых копий исходного сигнала ($x_j(n) = x(n) + \beta_0 E_1(w_j(n))$), определяется первый остаток $r_1(n) = M(x_j(n))$.

Шаг 2. Для $i = 1$ вычисляется первая мода: $IMF_1(n) = x(n) - r_1(n)$.

Шаг 3. Вычисляется второй остаток как усредненное локальное среднее значение шумовых копий первого остатка $r_1(n) + \beta_1 E_2(w_j(n))$ и определяется вторая мода:

$$\widetilde{IMF}_2(n) = r_1(n) - r_2(n) = r_1(n) - M(r_1(n) + \beta_1 E_2(w_j(n))).$$

Шаг 4. На последующих этапах для $i = 3, \dots, I$ вычисляется i -й остаток

$$r_i(n) = M(r_{i-1}(n) + \beta_{i-1} E_i(w_j(n))).$$

Шаг 5. Вычисляется i -я мода

$$\widetilde{IMF}_i(n) = r_{i-1}(n) - r_i(n).$$

Шаг 6. Переход к шагу 4 для следующего значения i .

Константы $\beta_i = \varepsilon_i \text{std}(r_i(n))$ выбираются таким образом, чтобы получить желаемое отношение сигнал/шум между добавленным шумом и остатком, к которому добавляется шум. Для МДЭМ отношение сигнал/шум добавленного шума и остатка увеличивается на порядок i , поскольку энергия шума в i -м остатке ($i > 1$) является лишь малой энергией шума, добавленного в начале алгоритма. С учетом этого установим β_0 так, чтобы ε_0 было прямо противоположно желаемому отношению сигнал/шум между первым добавленным шумом и анализируемым сигналом. Если мы выражаем отношение сигнал/шум как фактор стандартных отклонений, то $\beta_0 = \varepsilon_0 \text{std}(x(n)) / \text{std}(E_1(w_f(n)))$. Для получения шумовых реализаций с меньшими амплитудами для последних стадий декомпозиции в остальных модах будем воспринимать шум как результат его предварительной обработки ДЭМ, т.е. без их нормализации стандартным отклонением.

Оптимизация процесса декомпозиции на эмпирические моды

Из вышеизложенного следует, что для метода улучшенной ПМДЭМАШ необходима оптимизация параметров функционирования:

- стандартного отклонения амплитуды добавляемого белого шума – N_{std} ;
- числа реализаций декомпозиций (добавлений белого шума) – NR ;
- максимально допустимого количества просеивающих итераций (критерий останова) – $MaxIter$;
- отношения сигнал/шум между добавленным шумом и остатком, к которому добавляется шум – $SNRFlag$ (если значение равно 1, тогда отношение сигнал/шум увеличивается для каждого этапа декомпозиции; если равно 2, то отношение сигнал/шум будет одинаковым для всех этапов).

Предварительные исследования выявили, что значения данных параметров настройки влияют на частотно-избирательные свойства декомпозиции [17].

Для повышения точности измерения ЧОТ требуется улучшение частотно-избирательных свойств декомпозиции. Работа является развитием ранее опубликованных трудов авторов [18, 19] и выполнена при финансовой поддержке по гранту Президента РФ, проект «Исследование информативно-значимых параметров речевых сигналов: поиск уникально новых признаков естественно выраженных эмоций для повышения точности оценки психоэмоционального состояния операторов систем управления с высокой степенью ответственности» № СП-246.2018.5, 2018–2020 гг.

Оптимизация процесса декомпозиции на эмпирические моды

На рис. 1 структурно представлен предложенный авторами способ оптимизации улучшенной ПМДЭМАШ для задачи измерения ЧОТ. Серым цветом обозначены этапы обработки речевого сигнала, синим – этапы оптимизации, зеленым – числовые данные. Суть способа заключается в оптимальной настройке параметров функционирования улучшенной ПМДЭМАШ на основе результатов оценки точности измерения ЧОТ. Для упрощения на рис. 1 представлены не все этапы обработки речевого сигнала. Рассмотрим некоторые этапы способа подробнее.

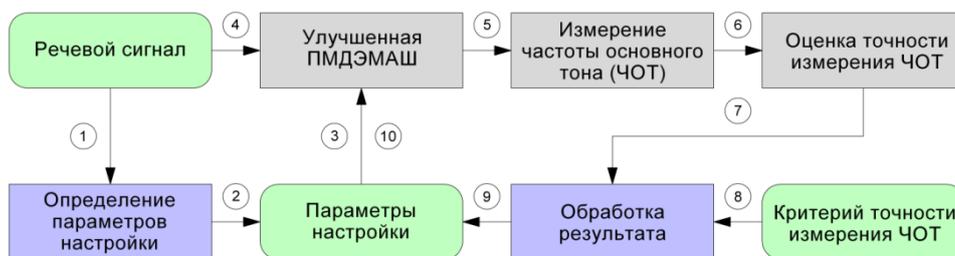


Рис. 1. Способ оптимизации улучшенной ПМДЭМАШ для задачи измерения ЧОТ

Определение параметров настройки заключается в вычислении исходного значения стандартного отклонения амплитуды речевого сигнала для дальнейшего вычисления коэффициента β_0 и установки диапазонов настройки параметров N_{std} , NR и $MaxIter$.

Разложение с помощью улучшенной ПМДЭМАШ обеспечивает [20]:

- адаптивную декомпозицию, так как базисные функции, используемые при разложении, извлекаются непосредственно из исходного речевого сигнала и позволяют учитывать только ему свойственные особенности (скрытые модуляции, области концентрации энергии и т.п.);
- минимальный уровень остаточного белого шума;
- отсутствие паразитных ЭМ, возникающих на ранних этапах декомпозиции вследствие перекрытия масштабно-энергетических пространств мод.

Измерение ЧОТ заключается в выполнении двух этапов: определение ЭМ, содержащей основной тон и непосредственное измерение частоты. В результате разложения речевой сигнал представлен набором ЭМ. Суть определения моды с основным тоном заключается в предположении, что информативные моды имеют большую энергию, чем неинформативные моды [17]. Для сжатия амплитуды сигнала ЭМ в большом динамическом диапазоне применяется логарифмирование энергии:

$$LE_i = \log_2 \sum_{n=1}^N (IMF_i(n))^2,$$

где LE_i – логарифм энергии ЭМ.

Процесс определения ЭМ с основным тоном заключается в последовательном вычислении разницы значений логарифмов энергии по модулю между текущей и последующей модами

$$d = |LE_i - LE_{i+1}|,$$

где d – разница между значениями логарифмов энергии между текущей и последующей ЭМ.

В результате из последовательности полученных значений d большему из них соответствует резкий спад энергии между информативной ЭМ, содержащей основной тон и неинформативной – трендовой [17–19].

Измерение ЧОТ осуществляется с использованием функции измерения мгновенной энергии сигнала – оператора Тигра, обладающего простотой, эффективностью и хорошей восприимчивостью к изменению амплитуды

$$T(n) = (IMF_{i,PF}(n))^2 - IMF_{i,PF}(n-1) \cdot IMF_{i,PF}(n+1),$$

где $T(n)$ – оператор Тигра; $IMF_{i,PF}(n)$ – ЭМ, содержащая основной тон.

Для измерения ЧОТ используются близкорасположенные максимум функции оператора Тигра, между которыми определяется разница в дискретных отсчетах времени, вычисляется период основного тона в секундах и частота в герцах

$$P_0 = \frac{T_{\max}(n+2) - T_{\max}(n)}{f_d}, \quad f_0 = \frac{1}{P_0},$$

где P_0 – основной тон; f_0 – ЧОТ; $T_{\max}(n)$, $T_{\max}(n+1)$ – максимумы функции оператора Тигра; f_d – частота дискретизации.

В качестве критериев оценки используются коэффициенты грубых и мелких ошибок [21]. Коэффициент грубых ошибок (*Gross Pitch Error – GPE*) – безразмерная величина, равная отношению числа фрагментов S_{GPE} с отклонением измеренного значения ЧОТ более чем на 20 % от истинного значения к общему числу фрагментов S_{f_0} , содержащих основной тон:

$$GPE = \frac{S_{GPE}}{S_{f_0}} \cdot 100 \%.$$

Усредненный коэффициент мелких ошибок (*Mean Fine Pitch Error – MFPE*) – безразмерная величина, равная среднему значению отношения разности истинного $f_{0,true}$ и измеренного значений f_0 частоты к истинному значению фрагментов, содержащих основной тон без грубых ошибок:

$$MFPE = \frac{1}{S_{FPE}} \sum_{s=1}^{S_{FPE}} \left| \frac{f_{0,true} - f_0}{f_{0,true}} \right| \times 100 \%,$$

где S_{FPE} – количество фрагментов, содержащих основной тон без грубых ошибок; s – номер фрагмента, содержащего основной тон.

Обработка результатов измерения ЧОТ заключается в сопоставлении полученных значений GPE и $MFPE$ со значениями установленных критериев точности $GPE_{crit.}$ и $MFPE_{crit.}$. Вычисляемая разница данных значений запоминается и осуществляется изменением параметров N_{std} , NR и $MaxIter$ в установленном диапазоне значений. Далее процесс продолжается по циклу, пока значения GPE и $MFPE$ не будут максимально близки к установленным критериям точности $GPE_{crit.}$ и $MFPE_{crit.}$.

Исследование способа оптимизации

Задачей исследования являлось определение оптимальных параметров настройки функционирования улучшенной ПМДЭМАШ для измерения ЧОТ. Исследования проводились с использованием 100 синтезированных мультгармонических сигналов, представляющих сумму нескольких гармонических составляющих основного тона с заранее известной частотой. Учитывая реальные условия, каждый тестовый сигнал подвергался частотному модулированию основного тона ν – изменению частоты в диапазоне 0–2,5 Гц/мс с шагом 0,5 Гц/мс. Исследования и анализ данных были выполнены в среде математического моделирования © Matlab (MathWorks).

В соответствии с методикой исследований параметры функционирования улучшенной ПМДЭМАШ подвергались настройке в следующих диапазонах:

- стандартное отклонение амплитуды добавляемого белого шума – от 0,01 до 0,1 мВ с шагом 0,01 мВ;
- количество реализаций декомпозиций – от 100 до 1000 раз с шагом 100;
- максимально допустимое количество просеивающих итераций – от 100 до 5000 раз с шагом 500;
- отношение сигнал/шум между добавленным шумом и остатком, к которому добавляется шум, если значение равно 1, тогда отношение сигнал/шум увеличивается для каждого этапа декомпозиции; если равно 2, то отношение сигнал/шум будет одинаковым для всех этапов.

Критерии точности измерения ЧОТ: $GPE_{crit.} = 20\%$ и $MFPE_{crit.} = 30\%$. В табл. 1–4 представлены результаты исследований, усредненные для 100 мультгармонических сигналов.

В табл. 1 представлены результаты исследований – зависимости значений GPE и $MFPE$ от N_{std} при первоначальных значениях $NR = 100$, $MaxIter = 100$, $SNRFlag = 1$ и различных значениях ν .

Таблица 1

Зависимость коэффициента грубых ошибок и усредненного коэффициента мелких ошибок от амплитуды добавляемого белого шума

N_{std} , мВ	ν , Гц/мс						Критерии оценки	
	0	0,5	1,0	1,5	2,0	2,5		
0,01	0	0	0	0	0	0	Значения, %	GPE
	1,82	6,04	7,12	8,44	10,76	11,87		$MFPE$
0,02	0	0	0	0	0	0		GPE
	1,77	5,56	6,91	8,11	10,21	11,56		$MFPE$
0,03	0	0	0	0	0	0		GPE
	1,68	5,31	7,03	8,01	10,43	11,98		$MFPE$
0,04	0	0	0	0	0	0		GPE
	1,46	4,25	5,98	6,81	9,12	10,96		$MFPE$
0,05	0	0	0	0	0	0		GPE
	1,44	4,53	6,11	6,97	8,82	11,20		$MFPE$
0,06	0	0	0	0	0	0	GPE	
	1,37	5,27	5,87	6,92	9,21	10,90	$MFPE$	
0,07	0	0	0	0	0	0	GPE	
	1,31	4,61	5,31	6,39	8,83	10,24	$MFPE$	
0,08	0	0	0	0	0	0	GPE	
	1,21	2,65	3,82	5,34	7,65	9,82	$MFPE$	
0,09	0	0	0	7,90	8,41	9,24	GPE	
	1,64	2,9	4,25	6,71	8,77	12,65	$MFPE$	
0,10	0	0	5,49	8,10	9,12	12,35	GPE	
	2,75	4,67	6,39	8,25	11,35	17,68	$MFPE$	

Примечание: Для малых значений $N_{std} = 0,01 \dots 0,08$ мВ $GPE = 0$.

В табл. 2 представлены результаты исследований – зависимость значений $MFPE$ от NR при оптимальном значении $N_{std} = 0,08$, первоначальных $MaxIter = 100$, $SNRFlag = 1$ и различных v . Учитывая, что при оптимальном значении $N_{std} = 0,08$ коэффициент грубых ошибок равен нулю, в дальнейшем исследования проводились без вычисления GPE .

Таблица 2

Зависимость усредненного коэффициента мелких ошибок от количества реализаций декомпозиций

NR	v , Гц/мс						Критерий оценки	
	0	0,5	1,0	1,5	2,0	2,5		
100	1,21	2,65	3,82	5,34	7,65	9,82	Значение, %	$MFPE$
200	1,20	2,54	3,77	5,12	7,36	9,45		
300	1,18	2,53	2,75	5,02	7,35	9,39		
400	1,17	2,51	3,72	4,96	7,32	9,34		
500	1,15	2,49	3,65	4,83	7,29	9,27		
600	1,13	2,47	3,62	4,79	7,21	9,19		
700	1,12	2,41	3,56	4,75	7,12	9,12		
800	1,11	2,34	3,45	4,67	7,07	9,05		
900	1,10	2,21	3,37	4,64	6,94	8,97		
1000	1,09	2,13	3,21	4,72	6,87	8,91		

В табл. 3 представлены результаты исследований – зависимость значений $MFPE$ от $MaxIter$ при оптимальных значениях $N_{std} = 0,08$, $NR = 1000$, первоначальном значении $SNRFlag = 1$ и различных v .

Таблица 3

Зависимость усредненного коэффициента мелких ошибок от максимально допустимого количества просеивающих итераций

$MaxIter$	v , Гц/мс						Критерии оценки	
	0	0,5	1,0	1,5	2,0	2,5		
100	1,09	2,13	3,21	4,72	6,87	8,91	Значение, %	$MFPE$
500	1,09	2,13	3,21	4,72	6,87	8,89		
1000	1,09	2,13	3,20	4,72	6,87	8,88		
1500	1,09	2,13	3,20	4,71	6,86	8,88		
2000	1,08	2,12	3,19	4,71	6,86	8,87		
2500	1,07	2,12	3,19	4,70	8,85	8,87		
3000	1,07	2,12	3,18	4,70	6,84	8,87		
3500	1,07	2,11	3,18	4,69	6,83	8,87		
4000	1,07	2,11	3,17	4,69	6,82	8,86		
4500	1,07	2,10	3,17	4,68	6,81	8,86		
5000	1,07	2,10	3,17	4,68	6,81	8,86		

В табл. 4 представлены результаты исследований – зависимость значений $MFPE$ от $SNRFlag$ при оптимальных значениях $N_{std} = 0,08$, $NR = 1000$, $MaxIter = 4500$ и различных v .

Таблица 4

Зависимость усредненного коэффициента мелких ошибок от отношения сигнал/шум между добавленным шумом и остатком

$SNRFlag$	v , Гц/мс						Критерии оценки, %	
	0	0,5	1,0	1,5	2,0	2,5		
1	1,07	2,10	3,17	4,68	6,81	8,86	Значение, %	$MFPE$
2	1,04	2,07	3,11	4,56	6,74	8,62		

На рис. 2 представлен пример разложения вокализованного участка реального речевого сигнала на ЭМ с использованием улучшенной ПМДЭМАШ при:

- начальных значениях диапазона параметров настройки ($N_{std} = 0,01$, $NR = 100$ и $MaxIter = 100$);
- конечных значениях диапазона параметров настройки ($N_{std} = 0,1$, $NR = 1000$ и $MaxIter = 5000$);
- оптимальных значениях параметров настройки ($N_{std} = 0,08$, $NR = 1000$ и $MaxIter = 4500$).

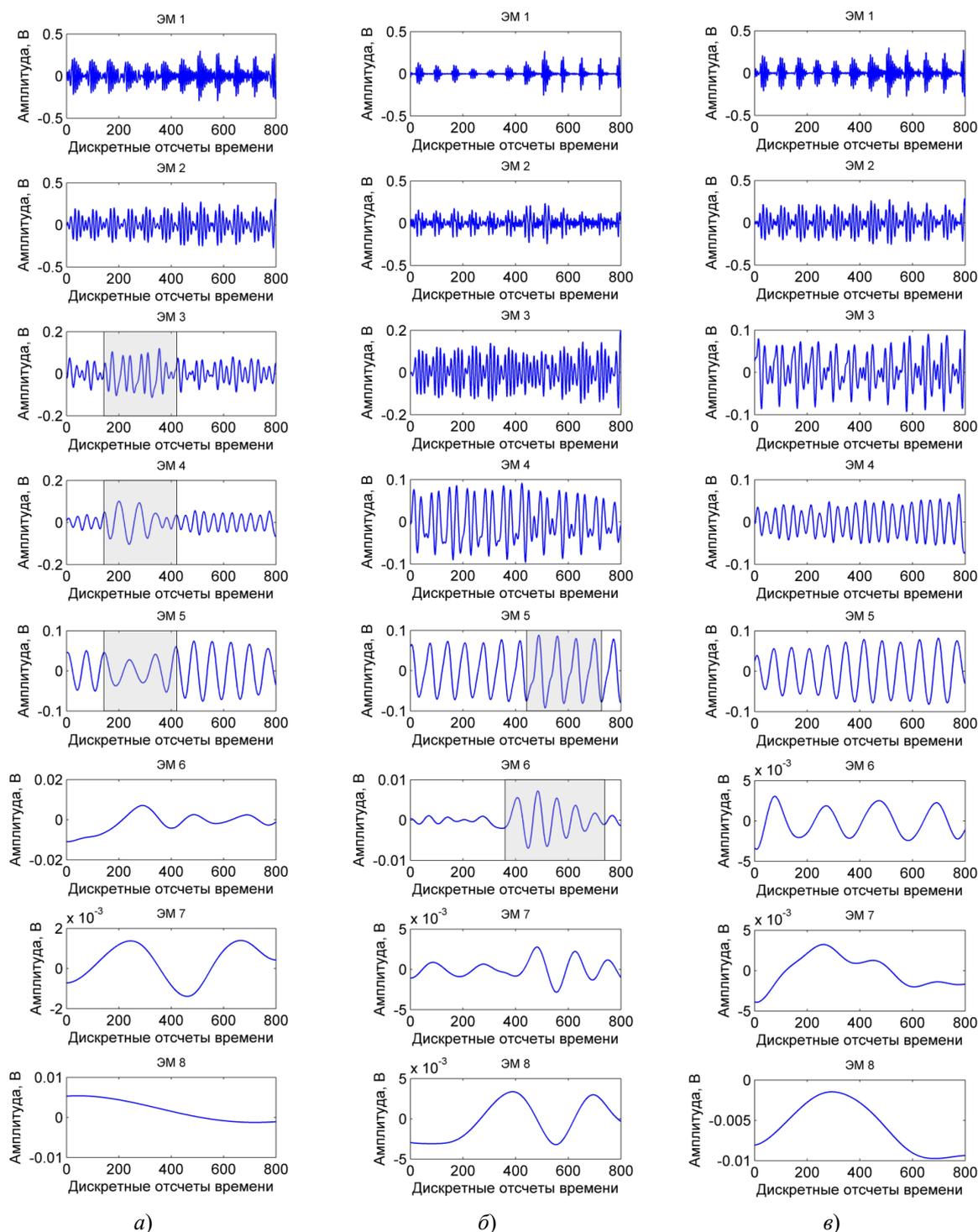


Рис. 2. Результаты разложения вокализованного участка реального речевого сигнала с использованием улучшенной ПМДЭМАШ: *а* – в начале диапазона; *б* – в конце диапазона; *в* – при оптимальных параметрах

Отношение сигнал/шум между добавленным шумом и остатком, к которому добавляется шум, будет одинаковым для каждого этапа декомпозиции, т.е. $SNRFlag = 2$.

На рис. 3 представлено спектральное распределение полученных мод, иллюстрирующее частотно-избирательные свойства метода улучшенной ПМДЭМАШ при начальных, конечных и оптимальных значениях параметров настройки.

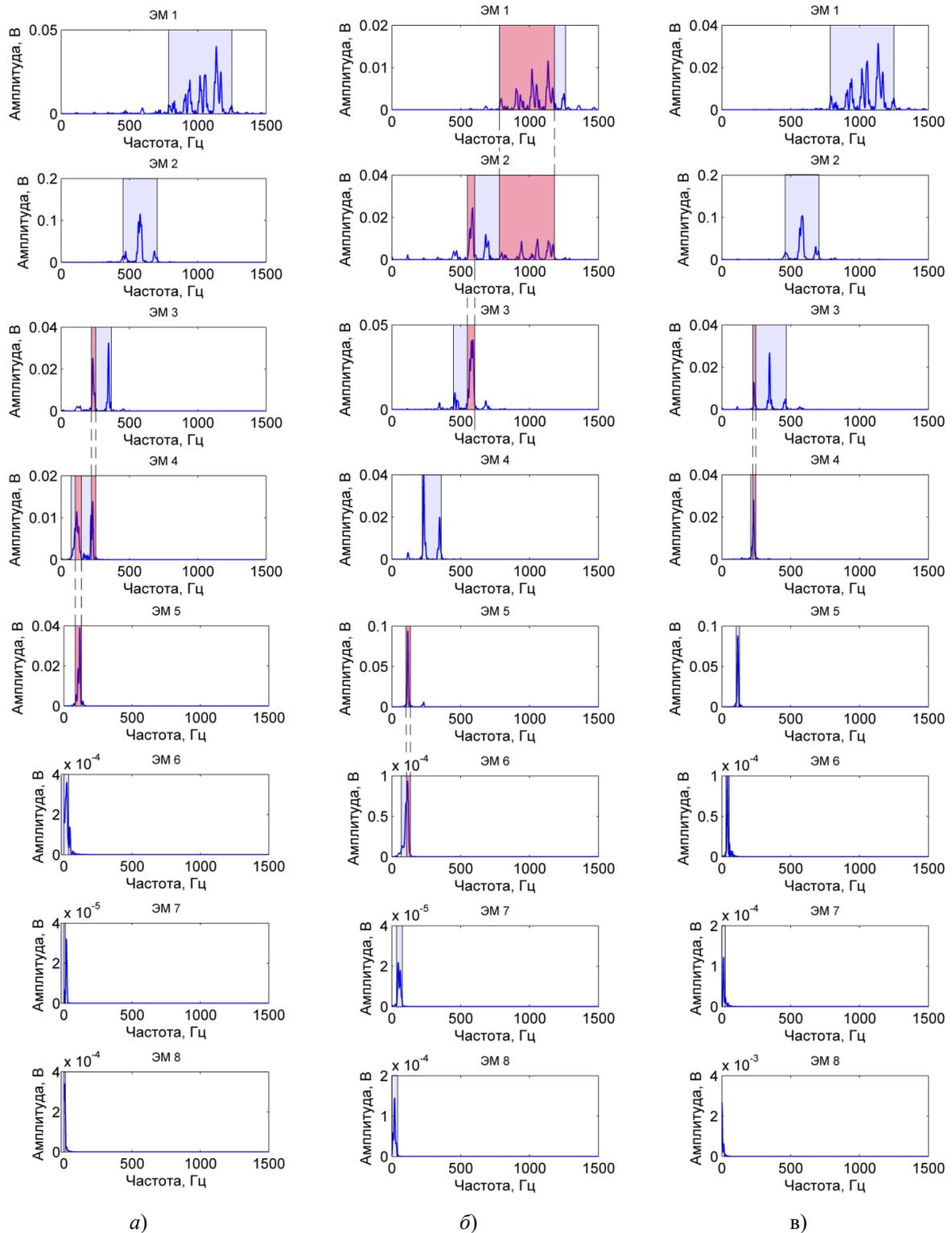


Рис. 3. Спектральное распределение полученных мод для улучшенной ПМДЭМАШ от параметров настройки:
 а – в начале диапазона; б – в конце диапазона; в – при оптимальных параметрах

Заключение

Из табл. 1 следует, что наилучшие результаты значений усредненного коэффициента мелких ошибок достигаются при стандартном отклонении амплитуды добавляемого белого шума 0,08 мВ. Причем данная закономерность прослеживается для всех значений скорости изменения ЧОТ. При увеличении амплитуды добавляемого шума процент мелких ошибок возрастает, и это объясняется чрезмерным зашумлением сигнала. При уменьшении шума усредненный коэффициент мелких ошибок также увеличивается. Это объясняется возникновением явления смешивания ЭМ, аналогично классической ДЭМ, так как при малых значениях шум не обеспечивает создание новых экстремумов функции сигнала. Что касается значений коэффициента грубых ошибок, то здесь предложенный способ управления начинает «ошибаться» при зашумлении сигнала белым шумом со стандартным отклонением амплитуды от 0,09 мВ и скоростью изменения ЧОТ от 1 Гц/мс.

Из табл. 2 следует, что при увеличении количества реализаций декомпозиций (добавлений белого шума) уменьшается усредненный коэффициент мелких ошибок для всех значений скорости изменения ЧОТ. Однако в этой ситуации нужно учесть время, затрачиваемое на каждую реализацию декомпозиции. Поэтому в зависимости от приоритетов поставленной задачи – уменьшение погрешности или увеличение быстродействия, необходимо определиться с конкретным значением количества реализаций.

По результатам проведенных исследований в табл. 3 неоднозначная картина получена при определении значения оптимального количества просеивающих итераций. Для каждого значения скорости изменения ЧОТ наилучший усредненный коэффициент мелких ошибок достигается при разном количестве просеивающих итераций. По этой причине рекомендуется определять оптимальное количество просеивающих итераций по верхней границе полученных значений усредненного коэффициента мелких ошибок. В данном случае – 4500 итераций.

Из табл. 4 следует, что при одинаковом отношении сигнал/шум между добавленным шумом и остатком на всех этапах декомпозиции достигается наилучший результат для всех значений скорости изменения ЧОТ.

Осциллограммы и спектральное распределение ЭМ на рис. 2 и 3 иллюстрируют, что полученные оптимальные значения параметров настройки улучшенной ПМДЭМАШ при использовании мультитармонических сигналов обеспечивают приемлемые результаты и для реальных речевых сигналов.

Из рис. 2,а,б следует, что вследствие перекрытия масштабно-энергетических пространств в начале и конце диапазона изменений значений параметров настройки наблюдается явление смешивания мод (выделено серым цветом). По этой же причине наблюдается перекрытие частотных полос ЭМ (см. рис. 3,а,б, выделено красным цветом). Полное отсутствие смешивания мод и минимальное перекрытие частотных полос наблюдаются при оптимальных значениях параметров настройки улучшенной ПМДЭМАШ (рис. 2,в и 3,в).

Таким образом, представленный в статье способ оптимизации процесса улучшенной ПМДЭМАШ решает задачу повышения точности измерения ЧОТ за счет улучшения частотно-избирательных свойств.

Библиографический список

1. Алимуратов, А. К. Исследование частотно-избирательных свойств методов декомпозиции на эмпирические моды для оценки частоты основного тона речевых сигналов / А. К. Алимуратов // Труды Московского физико-технического института. – 2015. – Т. 7, № 3. – С. 56–68.
2. Алимуратов, А. К. Исследование частотно-избирательных свойств методов декомпозиции речевых сигналов на эмпирические моды / А. К. Алимуратов // Модели, системы, сети в экономике, технике, природе и обществе. – 2015. – № 3 (15). – С. 97–109.
3. Huang, X. Spoken Language Processing. Guide to Algorithms and System Development / X. Huang, A. Acero, N.-W. Hon // Prentice Hall. – New Jersey, 2001. – 980 p.
4. Рабинер, Л. Р. Цифровая обработка речевых сигналов : пер. с англ. / Л. Р. Рабинер, Р. В. Шафер. – М. : Радио и связь, 1981. – 496 с.
5. Баскаков, С. И. Радиотехнические цепи и сигналы / С. И. Баскаков. – М. : Высш. шк., 2000. – 214 с.
6. Малла, С. Вейвлеты в обработке сигналов : пер. англ. / С. Малла. – М. : Мир, 2005. – 212 с.

7. *Bovbel, E. I.* Speech parameters vector based on arithmetic fourier transform / E. I. Bovbel, I. E. Kheidorov // European Signal Processing Conference (EUSIPCO 1998). – 1998. – P. 1–3.
8. *Rady* Speech Recognition System Based on Wavelet Transform and Artificial Neural Network / R. Rady Engy, H. Yahia Ashraf, A. El-Dahshan El-Sayed, El-Borey Hatem // Egyptian Computer Science Journal. – 2013. – Vol. 37, № 3. – P. 85–96.
9. *Speech Recognition by Wavelet Analysis* / N. Trivedi, V. Dr. Kumar, S. Singh, S. Ahuja, R. Chadha // International Journal of Computer Applications. – 2011. – Vol. 15, № 8. – P. 27–32.
10. *Фант, Г. К.* Акустическая теория речеобразования / Г. К. Фант ; пер. с англ. Л. А. Варшавского и В. И. Медведева ; науч. ред. В. С. Григорьева. – М. : Наука, 1964. – 284 с.
11. *Alimuradov, A. K.* Research of Frequency-Selective Properties of Empirical Mode Decomposition Methods for Speech Signals' Pitch Frequency Estimation / A. K. Alimuradov // International Conference on Engineering and Telecommunication (EnT) (Nov. 18–19, 2015, Russia, Moscow). – Moscow, 2015. – P. 77–79.
12. *Huang, N. E.* The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis / N. E. Huang, Sh. Zheng, R. L. Steven // Proc. R. Soc. Lond. – 1998. – Vol. A454. – P. 903–995.
13. *Zhaohua, Wu.* Ensemble empirical mode decomposition: A noise – assisted data analysis method / Wu Zhaohua, N. E. Huang // Advances in Adaptive Data Analysis. – 2009. – Vol. 1, № 1. – P. 1–41.
14. *Yeh J.-R.* Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method / J.-R. Yeh, J.-S. Shieh, N. E. Huang // Adv. Adapt. Data Anal. – 2010. – Vol. 2, № 2. – P. 135–156.
15. *Torres, M. E.* A complete Ensemble Empirical Mode decomposition with adaptive noise / M. E. Torres, M. A. Colominas, G. Schlotthauer, P. Flandrin // International Conference on Acoustics, Speech and Signal Processing (ICASSP-1). – Prague (CZ), 2011. – P. 4144–4147.
16. *Colominasa, M. A.* Improved complete ensemble EMD: A suitable tool for biomedical signal processing / Marcelo A. Colominasa, Gaston Schlotthauera, Maria E. Torres // Biomedical Signal Processing and Control. – 2014. – Vol. 14. – P. 19–29.
17. *Алимурадов, А. К.* Исследование частотно-избирательных свойств методов декомпозиции на эмпирические моды для оценки частоты основного тона речевых сигналов / А. К. Алимурадов // Труды Московского физико-технического института. – 2015. – Т. 7, № 3. – С. 56–68.
18. *Measurement of Speech Signal Patterns under Borderline Mental Disorders* / A. K. Alimuradov, A. Yu. Tychkov, A. V. Kuzmin, P. P. Churakov, A. V. Ageykin, G. V. Vishnevskaya // Proceedings of the 21st Conference of Open Innovations Association FRUCT (Nov. 6–10, 2017, Helsinki, Finland). – Helsinki : FRUCT Oy, 2017. – P. 26–33.
19. *Алимурадов, А. К.* Разработка алгоритма обработки речевых сигналов для определения информативно-значимых параметров пограничных психических расстройств / А. К. Алимурадов, А. Ю. Тычков, П. П. Чураков, А. В. Агейкин // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2017. – № 3 (47). – С. 4–25.
20. *Алимурадов, А. К.* Помехоустойчивая обработка речевых сигналов на основе комбинированной множественной декомпозиции на эмпирические моды / А. К. Алимурадов, Ю. С. Квитка, А. П. Зарецкий, А. П. Кулешов // Труды Московского физико-технического института. – 2016. – Т. 8, № 3. – С. 43–53.
21. *Alimuradov, A. K.* An Algorithm for Measurement of the Pitch Frequency of Speech Signals Based on Complementary Ensemble Decomposition Into Empirical Modes / A. K. Alimuradov // Measurement techniques. – 2017. – Vol. 59, № 12. – P. 1316–1323.

Алимурадов Алан Казанферович

кандидат технических наук,

директор студенческого научно-производственного
бизнес-инкубатора,

Пензенский государственный университет

(Россия, г. Пенза, ул. Красная, 40)

E-mail: alansapfir@yandex.ru

Alimuradov Alan Kazanferovich

candidate of technical sciences,

director of student research and production business
incubator,

Penza State University

(40 Krasnaya street, Penza, Russia)

Квитка Юрий Сергеевич

аспирант,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: alansapfir@yandex.ru

Kvitka Yury Sergeevich

postgraduate student,
Penza State University
(40 Krasnaya street, Penza, Russia)

Чураков Петр Павлович

доктор технических наук, профессор,
кафедра информационно-измерительной техники
и метрологии,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: churakov-pp@mail.ru

Churakov Petr Pavlovich

doctor of technical sciences, professor,
sub-department of Information
and measuring equipment and metrology,
Penza State University
(40 Krasnaya street, Penza, Russia)

Тычков Александр Юрьевич

кандидат технических наук, заместитель директора
научно-исследовательского института
фундаментальных и прикладных исследований,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: tychkov-a@mail.ru

Tychkov Aleksandr Yur'evich

candidate of technical sciences,
deputy director at the Research Institute
for Basic and Applied Studies,
Penza State University
(40 Krasnaya street, Penza, Russia)

УДК 004.934

Алимурадов, А. К.

Повышение точности измерения частоты основного тона на основе оптимизации процесса декомпозиции речевых сигналов на эмпирические моды / А. К. Алимурадов, Ю. С. Квитка, П. П. Чураков, А. Ю. Тычков // Измерение. Мониторинг. Управление. Контроль. – 2018. – № 4 (26). – С. 53–65. – DOI 10.21685/2307-5538-2018-4-9.