

## Адаптивный метод измерения частоты основного тона с использованием двухуровневой авторегрессионной модели речевого сигнала

Андрей Владимирович Савченко<sup>1</sup>, Владимир Васильевич Савченко<sup>2</sup>✉

<sup>1</sup> Национальный исследовательский университет «Высшая школа экономики», Нижний Новгород, Россия, [avsavchenko@hse.ru](mailto:avsavchenko@hse.ru), <https://orcid.org/0000-0001-6196-0564>

<sup>2</sup> Нижегородский государственный лингвистический университет, Нижний Новгород, Россия, [vvsavchenko@yandex.ru](mailto:vvsavchenko@yandex.ru)✉, <https://orcid.org/0000-0003-3045-3337>

**Аннотация.** Рассмотрена задача определения частоты основного тона речевого сигнала в условиях действия белого гауссова шума. Предложен метод измерений данной частоты, который учитывает периодическую структуру спектра мощности вокализованных фреймов речи и основан на принципе накопления энергии гармоник в частотной области. Для этого в алгоритм обработки речевого сигнала введена процедура выравнивания огибающей спектра мощности с использованием двухуровневой авторегрессионной модели наблюдений: в пределах одного периода основного тона и в интервале нескольких таких периодов. При этом предусмотрена адаптация порядка авторегрессии нижнего уровня под наблюдаемый фрейм. Рассмотрен пример практической реализации адаптивного метода на базе метода Берга. Основные достоинства адаптивного метода по сравнению с известными аналогами – высокое быстродействие и повышенная помехоустойчивость – подтверждены результатами проведенного натурного эксперимента. В результате применения адаптивного метода получен выигрыш в пороговых сигналах 5–10 дБ.

**Ключевые слова:** акустика речи, речевой сигнал, акустические измерения, авторегрессионная модель, проблема априорной неопределенности, адаптивный подход

**Финансирование:** Исследование выполнено при финансовой поддержке Российского научного фонда (проект № 20-71-10010).

**Для цитирования:** Савченко А. В., Савченко В. В. Адаптивный метод измерения частоты основного тона с использованием двухуровневой авторегрессионной модели речевого сигнала // Измерительная техника. 2022. № 6. С. 60–66. <https://doi.org/10.32446/0368-1025it.2022-6-60-66>

## An adaptive method for measuring the pitch frequency using a two-level autoregressive model of a speech signal

Andrey V. Savchenko<sup>1</sup>, Vladimir V. Savchenko<sup>2</sup>✉

<sup>1</sup> National Research University Higher School of Economics, Nizhny Novgorod, Russia, [avsavchenko@hse.ru](mailto:avsavchenko@hse.ru), <https://orcid.org/0000-0001-6196-0564>

<sup>2</sup> Nizhny Novgorod State Linguistic University, Nizhny Novgorod, Russia, [vvsavchenko@yandex.ru](mailto:vvsavchenko@yandex.ru)✉, <https://orcid.org/0000-0003-3045-3337>

**Abstract.** The problem of determining the fundamental frequency of a speech signal in the presence of white Gaussian noise is considered. A new method for its measurement is proposed, which takes into account the periodic structure of the power spectrum of voiced speech frames and is based on the principle of harmonic energy accumulation in the frequency domain. For this purpose, a procedure for equalizing its spectral envelope was introduced into the speech signal processing algorithm using a two-level autoregressive observations model: within one and several periods of the fundamental tone. At the same time, the adaptation of the order of autoregression of the lower level to the observed frame is implemented. An example of the practical implementation of the adaptive method based on the Berg method is considered. With the use of author's software, an experiment was set up and carried out, and experimental estimates of the effectiveness of the new method were obtained. It is shown that, due to its adaptation, a gain in threshold signals of 5–10 dB is achieved.

**Keywords:** speech acoustics, voice signal, acoustic measurements, autoregressive model, a priori uncertainty, adaptive approach

**Financial Support:** the work is supported by RSF (Russian Science Foundation) grant no. 20-71-10010.

**Введение.** На протяжении многих лет частота основного тона (ЧОТ) широко используется в системах автома-

тической обработки речи самого разного назначения: от передачи речевого сигнала по цифровым каналам связи

до речевой аналитики и искусственного интеллекта [1, 2]. Таким образом, измерение ЧОТ давно стало классической задачей в области речевых технологий, и существует множество разных подходов к её решению. Однако далеко не все проблемы в данной области исследований достаточно полно освещены в настоящий момент [3–5]. Прежде всего, не решён вопрос о потенциальной точности измерения ЧОТ при учёте множества случайных факторов [5–7]. Проблема обостряется условиями априорной неопределённости в отношении тонкой структуры речевого сигнала, фонового шума, а также малых выборок наблюдений на относительно коротких (длительностью  $\tau = 20 \dots 30$  мс) интервалах приближительной стационарности (квазистационарности) сигнала [8–10]. Для измерения ЧОТ с высокой точностью в этих условиях требуется адаптивный подход [11, 12]. Поэтому актуальной представляется тема настоящего исследования, основанного на принципе накопления энергии речевого сигнала в частотной области. При этом используется авторегрессионная (АР) модель двух уровней параметризации [13]: в расчёте на тонкую структуру спектра мощности речевого сигнала, с одной стороны, и на его спектральную огибающую, с другой. АР-модели первого и второго уровней имеют порядки  $p$  и  $q$  соответственно, причём  $p \gg 1$ ,  $q \ll p$ .

Цель исследования – повышение точности измерений ЧОТ речевого сигнала на фоне помех типа белого гауссова шума в режиме скользящего окна наблюдений конечной длительности  $\tau$ .

**Постановка задачи.** Согласно акустической теории речеобразования [1] спектральная плотность мощности (СПМ) вокализованных отрезков (фреймов) речевого сигнала  $x(t)$  имеет квазипериодическую структуру [8, 14]. Её минимальная частота  $F_0$  или период повторения последовательности  $K = 0,5F/F_0 \gg 1$  узкополосных (квазигармонических) составляющих, где  $F$  – частота временной дискретизации, и определяет понятие ЧОТ [5]. При этом оптимальная обработка речевого сигнала состоит из двух последовательных операций [15]: внутрипериодной автокомпенсации сигнала  $x(t)$  на входе измерителя ЧОТ и межпериодного накопления сигнала  $y(t)$  на выходе автокомпенсатора. При практической реализации используются предикторы речи (от англ. prediction – предсказание) двух разных уровней линейного предсказания речевого сигнала [13]: в пределах одного периода основного тона  $T_0 = F_0^{-1} \gg T = F^{-1}$  и на интервале нескольких таких периодов. Внутрипериодный предиктор (автокомпенсатор) играет роль фильтра выравнивания (spectrum flattening) [16] или обеления (whitening) [17, 18] огибающей СПМ речевого сигнала в полосе рабочих частот  $f \leq 0,5F$ . В идеале на выходе автокомпенсатора можно получить периодическую (с периодом  $T_0$ ) последовательность импульсов возбуждения голосового тракта диктора [16, 19]. Для выделения данной последовательности из сигнала  $y(t)$  на фоне шума предназначен второй, межпериодный предиктор в функции межпериодного накопителя речевого сигнала [5]. Он выполняется по схеме рекурсивного (гребенчатого) фильтра в многоканальном варианте – в расчёте на определённый диапазон возможных

(альтернативных) значений ЧОТ в процессе акустических измерений [21]. При этом наилучшую альтернативу условный наблюдатель выбирает по принципу максимизации мощности сигнала на выходе определённого канала межпериодного накопителя. Однако в общем случае это весьма затратное техническое решение, которое плохо сочетается с режимом измерений ЧОТ в реальном времени. Поэтому на практике в роли накопителя широко применяют простейший фильтр нижних частот [16] или близкую к нему в принципиальном отношении кепстральную обработку речевого сигнала [22]. При этом окончательный блок измерителя ЧОТ выполняется по схеме частотного детектора. Однако низкочастотная фильтрация неразрывно связана с частичными потерями полезной информации [23]. Поэтому авторами настоящей статьи предпринята попытка создания субоптимального метода измерения ЧОТ по наблюдаемому фрейму конечной длительности  $\tau$  с накоплением энергии всех  $K$  гармоник ЧОТ при использовании двухуровневой АР-модели речевого сигнала.

**Двухуровневая АР-модель.** Наиболее естественным способом обработки сигналов в частотной области является дискретное Фурье-преобразование [24]. Однако при учёте малой длительности речевого фрейма  $x(t)$  и в виду недостаточной разрешающей способности по частоте такой вариант мало пригоден для практического применения в рамках рассматриваемой задачи. Проблема преодолевается радикально путём применения высокоскоростных параметрических методов спектрального анализа, основанных на АР-модели гласных звуков речи [17]. В частотной области данная модель определяется выражением для СПМ общего вида

$$G_p(f) = \sigma_p^2 T \left| 1 - \sum_{i=0}^p a_p(i) \exp(-j2\pi i f T) \right|^{-2}; \quad (1)$$

$$f \leq 0,5F,$$

где  $\sigma_p^2$  – дисперсия погрешности линейного предсказания в роли масштабного множителя;  $a_p(i)$  –  $i$ -й коэффициент линейной авторегрессии  $p$ -го порядка.

Порядок  $p$  устанавливает наблюдатель исходя из поставленной задачи [11, 12]. Для моделирования тонкой структуры речевого сигнала порядок  $p \geq 2K$  ограничен снизу удвоенным числом гармоник ЧОТ. Например, в диапазоне  $F_0 = 80 \dots 200$  Гц значений ЧОТ дикторов-мужчин [14] при  $F = 8$  кГц будем иметь  $K = 4000/(80 \dots 200) = 20 \dots 50$  гармоник с частотами  $F_0, 2F_0, \dots, KF_0$ . В данном случае получаем  $p \geq 100$ . Напротив, для моделирования огибающей СПМ требуется существенно более низкий порядок  $q$ . В таком случае выражение (1) примет вид инверсного преобразования [13]:

$$G_q(f) = \sigma_q^2 T \left| 1 - \sum_{i=0}^q a_q(i) \exp(-j2\pi i f T) \right|^{-2} = |A_q(jf)|^{-2}; \quad (2)$$

$$f \leq 0,5F$$

квадрата модуля комплексного коэффициента передачи

$$A_q(jf) = (\sigma_q^2 T)^{-0,5} \left( 1 - \sum_{i=1}^q a_q(i) \exp(-j2\pi i f T) \right) \quad (3)$$

линейного трансверсального фильтра  $q$ -го порядка с константой  $(\sigma_q^2 T)^{-0.5}$  в качестве коэффициента усиления. Здесь  $a_q(i)$  –  $i$ -й элемент вектора  $\mathbf{a}_q$  АР-коэффициентов  $q$ -го порядка. СПМ сигнала  $y(t)$  на выходе фильтра (3) в частотной области описывается выражением

$$G_{p,q}(f) = |A_q(jf)|^2 G_p(f) = G_p(f) / G_q(f); \quad \left. \begin{array}{l} f \leq 0.5F. \end{array} \right\} \quad (4)$$

В идеальном случае – при точно подобранных под наблюдаемый фрейм  $x(t)$  порядках  $p, q$  авторегрессий двух разных уровней – выражение (3) определяет ключевой элемент в составе оптимального измерителя ЧОТ, а именно внутрипериодный обеляющий фильтр. В таком случае выражение (4) будет определять линейчатую СПМ последовательности импульсов возбуждения голосового тракта диктора. Измеряя период их повторения, получаем оптимальную оценку ЧОТ  $F_0$ . Проблема состоит, главным образом, в выборе оптимального значения порядка  $q$  АР-параметризации нижнего уровня, поскольку порядок  $p$  жёстко обусловлен числом  $K$  гармоник ЧОТ в рабочем диапазоне частот  $f \leq 0.5F$ . Решение данной проблемы в рамках адаптивного подхода приведено далее.

**Адаптивный метод.** Предположим, что обеляющий фильтр (3) оптимально адаптирован под наблюдаемый фрейм  $x(t)$ . В таком случае в соответствии с (4) СПМ сигнала  $y(t)$  будет состоять из  $K = 0.5F / F_0 \gg 1$  ярко выраженных гармонических составляющих с равномерным сдвигом между собой по частоте, в идеале – на  $F_0$ . Измерив значение сдвига тем или иным способом, можно получить искомую оценку ЧОТ  $\hat{F}_0$ . Задача состоит, таким образом, в выборе способа измерения.

Эффективный способ основан на вычислении автокорреляционной функции (АКФ) спектра мощности речевого сигнала [25]. Этому, как правило, предшествует амплитудное квантование СПМ (4) на два уровня:

$$\forall f \leq 0.5F: G_0(f) = \begin{cases} 1, & G_{p,q}(f) > g_0; \\ 0, & G_{p,q}(f) \leq g_0; \end{cases} \quad (5)$$

где  $g_0$  – порог квантования, пропорционально связанный со средней мощностью сигнала  $y(t)$  в частотной области.

На множестве дискретных частот  $f_m = f_{m-1} + \Delta f$ ,  $m = 1, 2, \dots, M$ ,  $M = 0.5F / \Delta f$ , запишем оценку ЧОТ в общем виде

$$\hat{F}_0 = \operatorname{Argmax}_{m_1 \leq r \leq m_2} Q_{p,q}(f_r) \triangleq \hat{F}_0(q), \quad (6)$$

где  $Q_{p,q}(f_r)$  – нормированная АКФ квантованного спектра (5).

Аргументом  $Q_{p,q}(f_r)$  является частотный сдвиг  $f_r = r \Delta f$  в границах  $m_1 \Delta f \leq f_r \leq m_2 \Delta f$  анализируемого диапазона частот. Эти границы регулируются выбором двух соответствующих констант  $m_1 \leq m_2 \leq 0.5F / \Delta f$ . Причём в целях исключения для правила (6) тривиального результата  $\hat{F}_0 = 0$  константа  $m_1 = \inf F_0 / \Delta f > 0$  не может быть установлена равной нулю. Более того, в расчёте на определённую категорию дикторов

она может намного превышать единицу. При этом амплитуда первого пика АКФ в области боковых лепестков

$$Q_{p,q}^* \triangleq Q_{p,q}(f_r) \big|_{f_r = \hat{F}_0(q)} \quad (7)$$

характеризует степень надёжности формируемой согласной (6) оценки ЧОТ. Чем ближе этот пик к единице, тем надёжнее соответствующая оценка  $\hat{F}_0(q)$ . Она существенным образом зависит от обоих порядков двухуровневой АР-модели (1). Однако эта зависимость носит неравноценный характер. Так, если значение  $p = 100$  обусловлено нижней границей ЧОТ [13], примерно равной 80 Гц для всех без исключения категорий дикторов, то порядок  $q$  не может быть установлен постоянным даже в расчёте на конкретного диктора, если учитывать отмеченную выше структурную вариативность его речи.

Проблема заключается в том, что в соответствии с назначением обеляющего фильтра его вектор коэффициентов в выражении (3) должен быть адаптирован под текущую (мгновенную [23]) АР-оценку СПМ (1) порядка  $p \gg q$  при условии, что размерность  $q$  этого вектора на данный момент ещё не определена. Указанное противоречие разрешает двухэтапная процедура адаптации. На первом этапе вектор  $\mathbf{a}_q$  АР-коэффициентов адаптируется по наблюдаемому фрейму  $x(t)$  с использованием метода Берга одновременно в нескольких ( $L > 1$ ) вариантах его порядка  $q = q_1, q_2, \dots, q_L$ . И далее из них выберем вариант

$$q^* = \operatorname{Argmax}_{p=\text{const}} Q_{p,q}^* \quad (8)$$

по критерию максимума пика (7). Данному варианту отвечает оценка ЧОТ вида

$$\hat{F}_0^* = \hat{F}_0(q) \big|_{q=q^*} = \hat{F}_0(q^*). \quad (9)$$

Выражения (1)–(9) в совокупности определяют адаптивный метод измерения ЧОТ с автоматическим выбором порядка  $q$  по результатам обработки наблюдаемого фрейма  $x(t)$ . Выражение (7) описывает в этом методе функцию межпериодного накопления, а выражения (8) и (9) – критерий максимума нормированной АКФ в частотной области.

Для практической реализации разработанного метода лучше всего подходят параметрические методы цифрового спектрального анализа, рассчитанные на применение в условиях малых выборок наблюдений, например, метод Берга, который гарантирует устойчивость формируемой АР-модели (2) в любом случае [24]. Ввиду высоких динамических свойств этот метод можно успешно использовать и для адаптации АР-оценки СПМ (1) высокого порядка  $p = 100$ . Именно такой вариант представляется наиболее эффективным техническим решением рассматриваемой задачи, поскольку спектральные оценки Берга разных порядков  $q_1 < q_2 < \dots < q_L < p$  формируются в рамках одной рекуррентной вычислительной процедуры нарастающего порядка  $p = 1, 2, \dots$ . Эффективность предложенного метода подтверждена данными проведенного эксперимента.

**Программа и результаты эксперимента.** На первой стадии экспериментального исследования объектом служили сигналы шести гласных звуков русской речи в их аддитивной смеси с фоновым шумом. При этом роль контрольного диктора выполнял один из авторов настоящей статьи, а источником шума служил генератор белого гауссова шума в программном исполнении. Дисперсия шума, а вслед за ней и отношение сигнал/шум (ОСШ) на входе измерителя ЧОТ варьировались в пределах  $h^2=0...40$  дБ. Длительность каждого сигнала (2,5–3 с) изначально рассчитана на его автоматическое разделение на множество коротких ( $\tau=30$  мс) фреймов  $x(t)$  при их частичном (по 20 мс) перекрытии во времени. При частоте дискретизации речевого сигнала  $F=8$  кГц размерность фреймов составила  $N=30 \cdot 8=240$  отсчётов. В результате по каждому гласному звуку речи была сформирована достаточно представительная база данных объёмом  $R=(2...3) \cdot 10^2=200...300$  независимых образцов для каждого отдельного значения ОСШ. Далее по каждому речевому фрейму  $x(t)$  методом Берга с шагом  $\Delta f=10$  Гц получены СПМ (1), (2). Порядок первой из них установлен равным  $p=100$ , порядок  $q=8...20$  варьировался в пределах заданного множества альтернатив. По результатам проведённых испытаний составлено множество альтернативных спектральных пар  $\{G_p(f); G_q(f)\}$ , с использованием которых вычислены соответствующие «обелённые» СПМ (4), в том числе в их бинарном варианте (5). Затем, следуя решающему правилу (6), получено множество альтернативных результатов  $\hat{F}_0(q)$  измерения ЧОТ в пределах каждого отдельного наблюдаемого фрейма  $x(t)$ . Из их числа в соответствии с критерием (9) отобраны оптимальные варианты оценки  $\hat{F}_0^*$  для каждого отдельного значения ОСШ. Инструментальная погрешность измерений  $0,5\Delta f$  сокращена при этом до 2–3 Гц путём интерполяции зависимости  $Q_{p,q}(f_r)$  в окрестностях её первого локального максимума (7). В дальнейшем полученные оценки  $\hat{F}_0^*$  в пределах  $R$ -выборки экспериментальных данных  $\{x(t)\}$  статистически усреднены. В итоге при доверительной вероятности 0,9 относительная погрешность выборочной (итоговой) оценки ЧОТ  $\varepsilon_F=1,65R^{-0,5}=1,65(250...300)^{-0,5}=9...10\%$  [26]. Все основные вычисления в ходе эксперимента были выполнены с использованием авторской компьютерной программы Phoneme Training<sup>1</sup>. Полученные результаты отражены в виде графиков и диаграмм.

На рис. 1 представлено типичное семейство СПМ сигналов гласного звука «а» контрольного диктора на выходах соответственно спектрального анализатора (1), обеляющего фильтра (4) и амплитудного спектрального квантователя (5) в условиях относительно невысокого ( $h^2=20$  дБ) фонового шума на входе. При сравнении графиков  $G_p(f)$ ,  $G_{p,q}(f)$  выявляются эффекты как

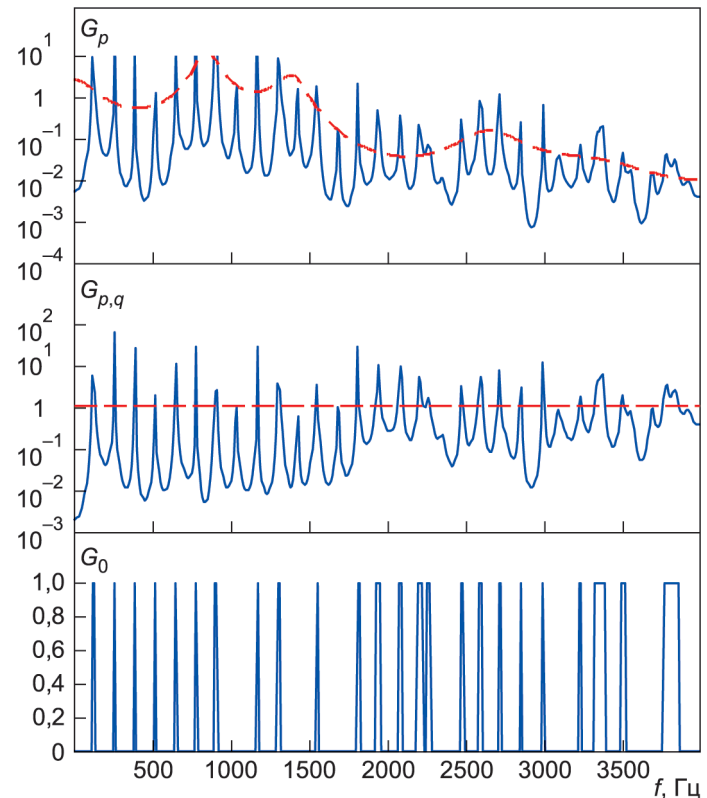


Рис. 1. Спектральная плотность мощности сигнала гласного звука «а» (сплошная линия) и её огибающая (штриховая линия) соответственно на входе измерителя ЧОТ  $G_p(f)$ , на выходе обеляющего фильтра  $G_{p,q}(f)$  и на выходе амплитудного квантователя  $G_0(f)$  при порядке  $q=10$

Fig. 1. The power spectral density of the signal of the vowel “a” (solid line) and its envelope (dashed line), respectively, at the input of the pitch frequency meter  $G_p(f)$ , at the output of the white filter  $G_{p,q}(f)$  and at the output of the amplitude quantizer  $G_0(f)$  with order  $q=10$

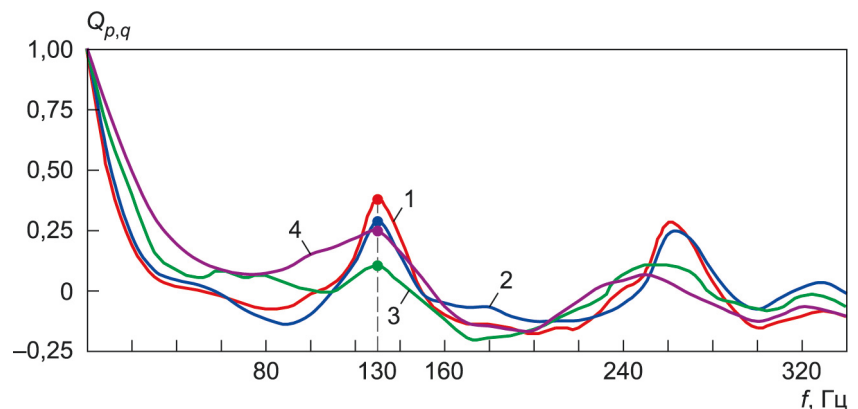


Рис. 2. Нормированная АКФ спектра мощности речевого сигнала на выходе амплитудного квантователя при отношении сигнал/шум 20 дБ (1, 2) и 0 дБ (3, 4) для порядков  $q$ , равных 12 (1), 10 (2, 3) и 8 (4)

Fig. 2. Normalized ACF of the speech signal's PSD at the output of the amplitude quantizer at SNR equal to 20 dB (1, 2) and 0 dB (3, 4) for cases  $q=12$  (1),  $q=10$  (2, 3) and  $q=8$  (4)

спектрального АР-моделирования, так и выравнивания спектральной огибающей. Горизонтальной штриховой линией на графике  $G_{p,q}(f)$  указан пороговый уровень

$$g_0 = M^{-1} \sum_{m=1}^M G_{p,q}(f_m),$$

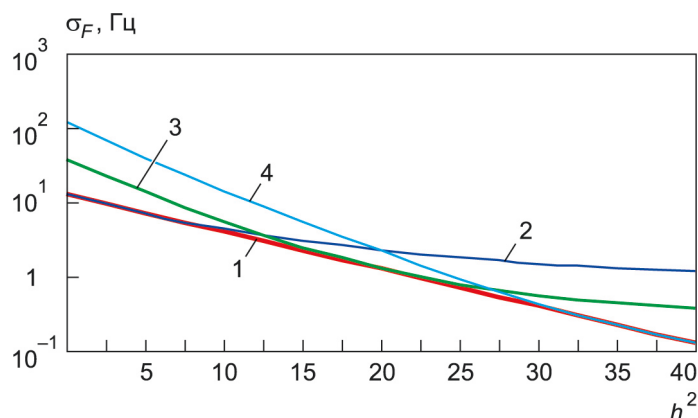


Рис. 3. Зависимость СКО  $\sigma_F$  погрешности измерения ЧОТ от отношения сигнал/шум  $h^2$  на входе измерителя для случаев  $q = q_{\text{опт}}$  (1),  $q = 8; 10; 15$  (2–4 соответственно)

Fig. 3. Dependence of RMS  $\sigma_F$  measurement error on SNR  $h^2$  at the pitch frequency meter input for cases  $q = q_{\text{опт}}$  (1),  $q = 8; 10; 15$  (2–4)

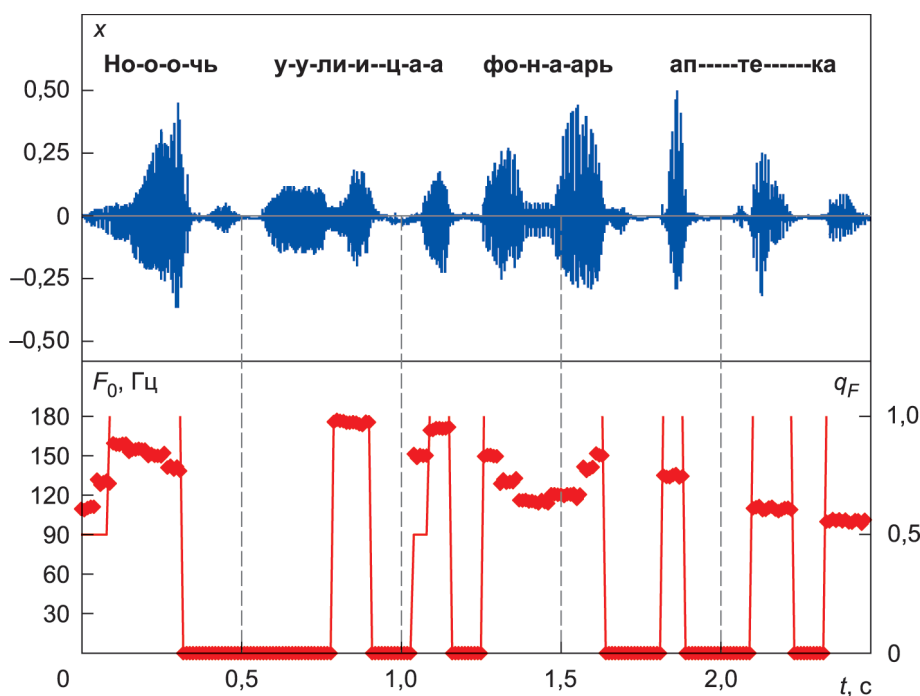


Рис. 4. Временная диаграмма контрольного речевого сигнала  $x(t)$ , траектория ЧОТ в динамике  $F_0(t)$  и диаграмма показателя надёжности  $q_F$  результатов

Fig. 4. Timing diagram of the control speech signal  $x(t)$ , the pitch frequency trajectory in dynamics  $F_0(t)$  and a diagram of the reliability index  $q_F$  of the results

определяемый в данном случае по формуле средней мощности сигнала  $y(t)$  в частотной области (4). График  $G_0(f)$  наглядно иллюстрирует назначение амплитудного квантователя в составе измерителя ЧОТ – нормализацию по форме и по уровню гармонических составляющих спектра мощности речевого сигнала.

Достигнутый при этом эффект в полной мере отражает рис. 2, на котором представлено семейство графиков нормированной АКФ спектра мощности на выходе амплитудного квантователя (5) при ОСШ, равном 20 и 0 дБ для порядков АР-модели  $q = 12; 10; 8$ . Амплитуда (7) первого пика АКФ (показана точкой) существенно зависит не только от ОСШ, но и от порядка  $q$  АР-модели. Однако положение этого пика в частотной области, а значит и значение  $\hat{F}_0 \approx 130$  Гц итоговой оценки (9) остаются стабильными во всех рассматриваемых случаях. Это является признаком высокой помехоустойчивости предложенного измерителя ЧОТ и, одновременно, стимулом для адаптации порядка  $q$  в рамках критерия (8).

В развитие сделанных выводов на рис. 3 представлены графики зависимости среднего квадратического отклонения (СКО)  $\sigma_F$  оценки ЧОТ сигнала гласного звука речи «а» от ОСШ (кривая 1) при использовании оптимального согласно критерию (8), (9) значения  $q_{\text{опт}}$  и других значениях порядка  $q$  (кривые 2–4). В области рабочих значений ОСШ  $h^2 = 10 \dots 20$  дБ кривые 2–4 существенно отличаются в худшую сторону от оптимального варианта.

Таким образом, вследствие адаптации измерителя ЧОТ к речевому фрейму  $x(t)$  на входе достигается выигрыш в пороговых сигналах порядка 5–10 дБ.

Сделанные выводы справедливы и для всех остальных гласных звуков речи контрольного диктора. Однако их оценки ЧОТ могут сильно различаться. Например, для звука «и» оценка  $\hat{F}_0$  составила около 120 Гц. С указанной точки зрения не только теоретическое, но и практическое значение имеют результаты второй, заключительной стадии проведённого эксперимента. Его объектом служила строка «Ночь, улица, фонарь, аптека» одноимённого стихотворения А. Блока. Аудиозапись этой строки сделана контрольным диктором в условиях рабочего кабинета при ОСШ порядка 20 дБ. В процессе автоматической обработки она разделена на последовательность фреймов  $x(t)$  длительностью  $\tau = 30$  мс каждый. Согласно методу (1)–(7) для каждого фрейма получена текущая оценка ЧОТ, временные диаграммы показаны на рис. 4. Синхронно с речевым сигналом  $x(t)$  (вверху) представлена последовательность формируемых согласно (9) оценок ЧОТ  $F_0(t)$  (внизу), а также диаграмма показателя их надёжности  $q_F$ .

определяемого по принципу: чем больше амплитуда бокового пика АКФ, тем выше надёжность результата измерений [16]:

$$q_F = \begin{cases} 1, & Q_{p,q}^* \geq 0,25; \\ 0,5, & 0,2 < Q_{p,q}^* \leq 0,25; \\ 0, & Q_{p,q}^* \leq 0,2. \end{cases} \quad (10)$$

Из рис. 4 следует, что оценки ЧОТ сильно разнятся между собой, значение оценки зависит не только от типа очередной фонемы, но и от интонации диктора в момент её произнесения. При этом разрывы и пропуски на временной диаграмме (см. рис. 4, внизу) указывают на соответствующие недостаточно вокализованные, в частности, назальные звуки речи диктора.

**Обсуждение полученных результатов.** Корреляционная обработка речевого сигнала давно и широко используется исследователями в целях выявления периодической составляющей в его тонкой структуре, правда, как правило, во временной области [5, 21]. Однако достигаемый при этом эффект сильно зависит от длительности  $\tau$  речевого фрейма. При  $\tau < 2T_0 = 2/F_0$  эффект равен нулю [9], но пропорционально возрастает при увеличении относительной длительности фрейма данных  $n_1 = \tau/T_0 = \tau F_0$ . Например, при  $F_0 = 100$  Гц и  $\tau = 30$  мс будем иметь  $n_1 = 3$ , но этого может оказаться недостаточно в условиях интенсивного фонового шума. Отсюда возникает острая проблема помехоустойчивости измерителей ЧОТ при работе в режиме реального времени [6, 17]. Картина резко изменяется при вычислении спектральной АКФ  $Q_{p,q}(f)$ . Здесь основным источником информации служит последовательность гармоник СПМ  $G_{p,q}(f)$  с периодом  $F_0$ . Соответственно, относительная длина интервала наблюдений в частотной области составляет  $n_2 = 0,5F/F_0 = K$ . Ввиду того, что  $F_0 \ll F$ , параметр  $K \gg 1$  (см. рис. 2), поэтому на значение  $n_2$  шум влияет в меньшей степени по сравнению с влиянием на  $n_1$ . А это признак более высокой помехоустойчивости.

Существенное для практики преимущество предложенного метода измерения ЧОТ состоит в автоматической оценке степени надёжности принимаемых решений (9) по уровню бокового пика АКФ (7) согласно критерию (8). Напротив, для большинства известных методов измерения ЧОТ процедура классификации такого рода является одной из основных составляющих их вычислительной сложности [27].

**Заключение.** В результате проведённого исследования предложен адаптивный метод измерения ЧОТ речевого сигнала в режиме скользящего окна конечной длительности  $\tau$ . Ключевое звено метода – обеляющий фильтр спектральной огибающей – адаптируется к наблюдаемому фрейму методом перебора альтернатив по критерию (8) максимума амплитуды боковых лепестков спектральной АКФ. В результате такой адаптации достигается выигрыш в пороговых сигналах 5–10 дБ. При этом вычислительные затраты на реализацию адаптивного метода порядка  $pN + LM$  элементарных операций умножения/деления отвечают производительности современных информационных систем при работе в реальном времени.

Полученные результаты предназначены для применения при разработке новых систем и технологий автоматической

обработки речи различного назначения, включая речевую связь [8, 13], прикладную акустику [4, 5] и другие, в целях отслеживания в динамике изменений эмоционального состояния их пользователей [26, 28].

#### Список источников / References

1. Rabiner L. R., Shafer R. W., *Theory and Applications of Digital Speech Processing*, Boston, Pearson, 2010, 1060 p.
2. Schenkman B. N., Gidla V. K., *Applied Acoustics*, 2020, vol. 163, 107214. <https://doi.org/10.1016/j.apacoust.2020.107214>
3. Souza G. V., Duarte J. M., Viegas F. et al, *Journal of Voice*, 2020, vol. 34, no. 4, pp. 641–648. <https://doi.org/10.1016/j.jvoice.2018.12.007>
4. Smith S. R., *The Journal of the Acoustical Society of America*, 2021, vol. 150, A113. <https://doi.org/10.1121/10.0007806>
5. Савченко А. В., Савченко В. В. Метод измерения частоты основного тона речевого сигнала для систем акустического анализа речи // Измерительная техника. 2019. № 3. С. 59–63. <https://doi.org/10.32446/0368-1025it.2019-3-59-63> [Savchenko A. V., Savchenko V. V., *Measurement Techniques*, 2019, vol. 62, no. 3, pp. 282–288. <https://doi.org/10.1007/s11018-019-01617-x>].
6. Yadav I. C., Shahnawazuddin S., Pradhan G., *Digital Signal Processing*, 2019, vol. 86, pp. 55–64. <https://doi.org/10.1016/j.dsp.2018.12.013>
7. Savchenko V. V., *Radioelectron. Commun. Syst.*, 2020, vol. 63, pp. 532–542. <https://doi.org/10.3103/S0735272720100039>
8. Gibson J. D., *Information*, 2016, vol. 32, no. 7. <https://doi.org/10.3390/info7020032>
9. Gu Yu., Wei H. L., *Information Sciences*, 2018, vol. 451–452, pp. 195–209. <https://doi.org/10.1016/j.ins.2018.04.007>
10. Cui S., Li E., Kang X., *IEEE International Conference on Multimedia and Expo (ICME)*, London, United Kingdom, 2020, pp. 1–6. <https://doi.org/10.1109/ICME46284.2020.9102765>
11. Savchenko V. V., Savchenko A. V., *Journal of Communications Technology and Electronics*, 2020, vol. 65, no. 11, pp. 1311–1317. <https://doi.org/10.1134/S1064226920110157>
12. Savchenko V. V., Savchenko A. V., *Radioelectron. Commun. Syst.*, 2019, vol. 62, no. 5, pp. 276–286. <https://doi.org/10.3103/S0735272719050042>
13. Savchenko V. V., *Radioelectron. Commun. Syst.*, 2021, vol. 64, no. 11, pp. 592–603. <https://doi.org/10.3103/S0735272721110030>
14. Kashani H. B., Sayadiyan A., *Computer Speech & Language*, 2018, vol. 50, pp. 105–125. <https://doi.org/10.1016/j.csl.2017.12.008>
15. Gibson J., *Information*, 2019, vol. 179, no. 10. <https://doi.org/10.3390/info10050179>
16. Markel J. D., Gray A. H., *Fundamental Frequency Estimation. In: Linear Prediction of Speech. Communication and Cybernetics*, Springer, Berlin, Heidelberg, 1976, vol. 12. [https://doi.org/10.1007/978-3-642-66286-7\\_8](https://doi.org/10.1007/978-3-642-66286-7_8)
17. Esfandiari M., Vorobyov S. A., Karimi M., *Signal Processing*, 2020, vol. 171, 107480. <https://doi.org/10.1016/j.sigpro.2020.107480>

18. Jaramillo A. E., Nielsen J. K., Christensen M. G., *27th European Signal Processing Conference (EUSIPCO)*, 2019, pp. 1–5. <https://doi.org/10.23919/EUSIPCO.2019.8902763>
19. Palaparthi A., Titze I. R., *Speech Communication*, 2020, vol. 123, pp. 98–108. <https://doi.org/10.1016/j.specom.2020.07.003>
20. Ширман Я. Д. и др. Радиозлектронные системы. Основы построения и теория: Справочник. Под ред. Я. Д. Ширмана. Изд. 2-е, перераб. и доп. М.: Радиотехника, 2007. 657 с. [*Radioelektronnyye sistemy. Osnovy postroeniya i teoriya. Spravochnik*, Ed. Ya. D. Shirman, 2nd ed., Moscow, Radiotekhnika Publ., 2007, 657 c.].
21. Sinha R., Shahnawazuddin S., *Computer Speech & Language*, 2018, vol. 48, pp. 103–121. <https://doi.org/10.1016/j.csl.2017.10.007>
22. Oppenheim A. V., Schaffer R. W., *IEEE Signal Processing Magazine*, 2004, vol. 21, no. 5, pp. 95–106. <https://doi.org/10.1109/MSP.2004.1328092>
23. Parlak C., Altun Yu., *Mathematical Problems in Engineering*, 2021, vol. 2021, 6658951. <https://doi.org/10.1155/2021/6658951>
24. Marple S. L., *Digital spectral analysis with applications*, 2nd ed., Mineola, New York, Dover Publications, 2019.
25. Levkov D. G., Panin A. G., Tkachev I. I., arXiv:2010.15145v4 [astro-ph.HE], 2021. <https://arxiv.org/abs/2010.15145>
26. Савченко Л. В., Савченко А. В. Метод измерений показателя динамики эмоционального состояния пользователя по его речевому сигналу в режиме реального времени // Измерительная техника. 2021. № 4. С. 72–79. <https://doi.org/10.32446/0368-1025it.2021-4-49-57> [Savchenko L. V., Savchenko A. V., *Measurement Techniques*, 2021, vol. 64, no. 12. <https://doi.org/10.1007/s11018-021-01935-z>].
27. Deng B., Jouvett D., Laprie Y., Steiner I. and Sini A., *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 5605–5609. <https://doi.org/10.1109/ICASSP.2017.7953229>
28. Akçay M. B., Oğuz K., *Speech Communication*, 2020, vol. 116, pp. 56–76. <https://doi.org/10.1016/j.specom.2019.12.001>

Статья поступила в редакцию 14.03.2022; одобрена после рецензирования 11.05.2022; принята к публикации 13.05.2022.  
The article was submitted 14.03.2022; approved after reviewing 11.05.2022; accepted for publication 13.05.2022.