Оценка мгновенной частоты основного тона речевого сигнала на основе многоскоростной обработки

Максим Иосифович Вашкевич,
кандидат технических наук, доцент Белорусского государственного университета информатики и радиоэлектроники (БГУИР)

Илья Сергеевич Азаров,
доктор технических наук, доцент БГУИР

Александр Александрович Петровский,
доктор технических наук, профессор кафедры электронных вычислительных средств БГУИР

Аннотация

В работе предлагается алгоритм оценки частоты основного тона, основанный на представлении речевого сигнала синусоидальной моделью с мгновенными параметрами. Алгоритмом предусмотрена следующая последовательность шагов: 1) декомпозиция сигнала на субполосные составляющие; 2) определение мгновенных параметров синусоидальной модели субполярных сигналов; 3) вычисление функции формирования кандидатов периода основного тона; 4) поиск локального контура частоты основного тона. Особенностью алгоритма является то, что ширина полос пропускания фильтров, используемых для декомпозиции, а также длительность кадра анализа масштабируются для каждого кандидата периода основного тона путем передискретизации сигнала. В работе делается сравнение предлагаемого алгоритма с широко используемыми оценщиками частоты основного тона RAPT, YIN, SWIPE', IRAPT и PEFAC. Предлагаемый алгоритм демонстрирует хорошее частотное и временное разрешение для сигналов, имеющих значительную частотную модуляцию, и показывает хорошую производительность как для чистых, так и для зашумленных сигналов.

Ключевые слова: частота основного тона, многоскоростная обработка

ВВЕДЕНИЕ

Надежное определение частоты основного тона требуется во многих приложениях обработки речи. В большинстве параметрических моделей, применяемых при кодировании, преобразовании и синтезе речевых сигналов, требуется оценка вокализованности/невокализованности и значение частоты основного тона. Использование алгоритма определения частоты основного тона с хорошим временным разрешени-

ем особенно необходимо для анализа/синтеза нестационарных звуков, которые обычно происходят на границах вокализованных сегментов и в моменты переходов. В то же время точная оценка контура частоты основного тона имеет большое значение при частотном анализе речевых сигналов, синхронизированном с частотой основного тона [1]. Понятие мгновенной частоты может быть естественным образом применено к частоте основного тона, если предположить, что речевой сигнал описывается гармонической моделью [2, 3]. Контур мгновенной частоты основного тона может быть извлечен из вокализованных участков речи, если рассматривать их как непрерывный и нестационарный процесс [4, 5].

Идея точной оценки частоты основного тона, разработанная в [6, 7], основана на декомпозиции сигнала на узкополосные компоненты и использовании их мгновенных частот в качестве исходных данных. Данный подход нашел применение в анализе речи и певческого голоса [8, 9], а также при создании устойчивого к ошибкам оценщика основного тона в [10]. Тем не менее у данного подхода есть фундаментальные ограничения, связанные с принципом неопределенности: нельзя достичь одинаково высокого разрешения для всего диапазона частот основного тона, используя банк фильтров с фиксированными параметрами. Для частотно-временного анализа в случае низкого голоса лучше использовать кадры большой длины, а для высоких голосов предпочтительно иметь короткую длину кадра и более широкие полосы у фильтров анализа. Компромиссное решение было найдено в [7], где использовались кадры длительностью 50 мс и фильтры с шириной полосы в 70 Гц, которое позволяло довольно точно оценить основной тон на женских голосах, но, как оказалось, подвержено грубым ошибкам на низких мужских голосах [11].

В данной статье описан алгоритм тонкой оценки частоты основного тона, осно-ванный на многоскоростной схеме анализа сигнала. Главной идеей является достиже-ние улучшения в точности оценки за счет подстройки параметров банка анализирующих фильтров для каждого кандидата периода основного тона. Предлагаемый алгоритм является эффективным для анализа как низких, так и высоких голосов. В алгоритме также предполагается, что вариация частоты основного тона пропорциональна её текущему значению. Увеличение ширины полосы анализирующих фильтров, необходимое для коротких кандидатов периода, приводит к смешиванию гармоник в субполосных сигналах при анализе низких голосов. Для компенсации этого эффекта предлагается использовать специального вида функцию формирования кандидатов периода (ФФКП) основного тона, которая менее чувствительна к смешению гармоник. В работе приводится как теоретическое обоснование предлагаемого алгоритма, так и практические аспекты его реализации. В заключительной части статьи оценивается производительность алгоритма на чистых и зашумленных сигналах.

1. МНОГОСКОРОСТНАЯ СХЕМА ВЫЧИСЛЕНИЯ ФУНКЦИИ ФОРМИРОВАНИЯ КАНДИДАТОВ ПЕРИОДА ОСНОВНОГО ТОНА

Предлагаемый оценщик частоты основного тона основан на синусоидальной модели, которая представляет детерминированную часть сигнала в виде суммы периодических компонент с нестационарными параметрами:

$$s(n) = \sum_{k=1}^{K} A_k(n)\cos(\varphi_k(n)) + r(n), \qquad (1)$$

где $\varphi_k(n) = \sum_{i=1}^{n} \omega_k(n) + \varphi_k(0)$, K — количество периодических компонент и $r(n)$ — шумовая компонента. Параметры данной модели (мгновенная амплитуда $A_k(n)$ и частота $\omega_k(n)$ в рад/отсчет) используются в качестве начальных данных для оценки частоты основного тона. Для получения параметров модели сигнал $s(n)$ раскладывается на комплексные субполосные составляющие ДПФ-модулированным банком фильтров, краткое описание которого приводится далее.

Равномерная сетка частот, соответствующая центральным частотам банка фильтров анализа, определяется как $k\omega_{step}$, $k = 1,2, ..., K$, $K\ \pi/\omega_{step}$, где ω_{step} — шаг по частоте в рад/отсчет. Импульсная характеристика k-го анализирующего фильтра определяется выражением:

$$h_k(n) = 2\frac{\sin(\omega_{bw} n)}{\pi n} w(n) e^{jkn\omega_{step}}, \quad (2)$$

где w_{bw} — половина ширины полосы пропускания фильтра и $w(n)$ — четная оконная функция.

Выход каждого канала банка фильтров является аналитическим сигналом $S_k(n)$ с ограниченной полосой, который можно представить как свертку входного сигнала $s(n)$ с импульсной характеристикой:

$$S_k(n) = \sum_{i=-\infty}^{\infty} h_k(i) s(n-i) = \mathrm{Re}(S_k(n)) + j\,\mathrm{Im}(S_k(n)). \quad (3)$$

Мгновенные параметры субполосных компонент могут быть получены следующим образом:

$$A_k(n) = \sqrt{\mathrm{Re}(S_k(n))^2 + \mathrm{Im}(S_k(n))^2}, \quad (4)$$

$$\varphi_k(n) = \arctan\left(\frac{-\mathrm{Im}(S_k(n))}{\mathrm{Re}(S_k(n))}\right), \omega_k(n) = \varphi_k'(n). \quad (5)$$

Чтобы избежать точек разрывов, в функции (5) применялась процедура развертывания фазы (*phase unwrapping*).

Мгновенные параметры $A_k(n)$, $\varphi_k(n)$ используются в качестве начальных данных для вычисления функции формирования кандидатов периода основного тона. Учитывая предположение, что вариация частоты основного тона пропорциональна её текущему значению, параметры анализирующего банка фильтров должны быть масштабированы для каждого кандидата периода следующим образом:

$$w_{step}(w_0) = w_0,\ w_{bw}(w_0) = \alpha w_0, \quad (6)$$

где w_0 — частота кандидата в рад/отсчет и α — допустимая относительная вариация тона. Длительность кадра анализа (N) должна быть подобрана, чтобы включать целое число периодов основного тона:

$$N = 2\pi L / w_0, \quad (7)$$

где L — число периодов в кадре анализа. Данная идея иллюстрируется на рисунке 1.

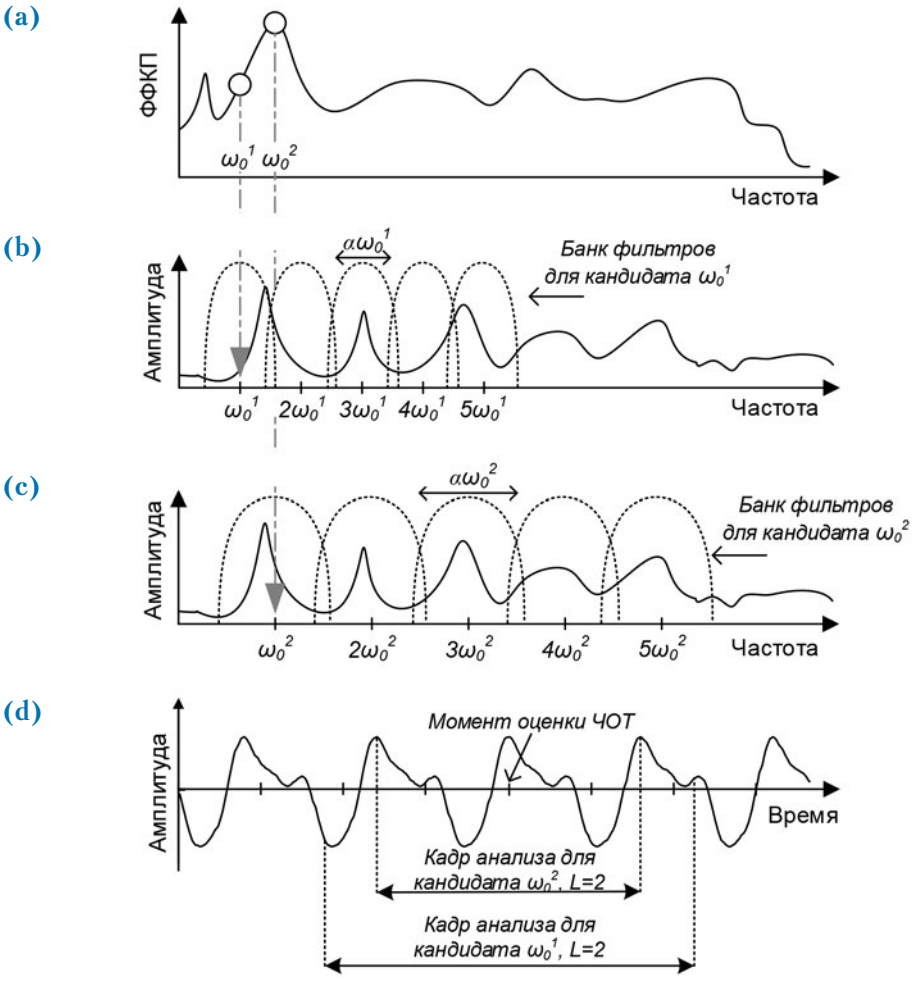

Рис. 1. Масштабирование анализирующего банка фильтров для каждого: (a) — функция формирования кандидата периода (ФФКП), (b) — амплитудный спектр и банк фильтров для кандидата ω_0^1, (c) — амплитудный спектр и банк фильтров для кандидата ω_0^2, (d) — исходный сигнал

Изменение параметров банка фильтров для каждого кандидата периода в общем случае является вычислительно затратным процессом. Альтернативой данному подходу может служить применение банка фильтров с фиксированными параметрами, но к сигналу с измененяемой частотой дискретизации. В этом случае можно определить частоту F_s дискретизации, кратной частоте кандидата периода (f_0):

$$F_s = Rf_0, \qquad (8)$$

где f_0 — частота в Гц и R — целое число. Используя (8) и учитывая, что $f_0 = w_0 F_s / (2\pi)$, выражение (7) принимает вид фиксированного по длительности кадра анализа для всех кандидатов на период:

$$N = RL. \qquad (9)$$

Параметр R определяет число гармоник, которые оставляются в передискретизированной версии сигнала:

$$K = \begin{cases} \dfrac{R-1}{2}, \text{ для нечетных } R \\ \dfrac{R}{2}-1, \text{ для четных } R. \end{cases} \quad (10)$$

Поскольку практическую значимость для определения частоты основного тона имеют лишь несколько первых гармоник, то для анализа можно использовать очень короткие по длительности кадры. Учитывая (8), выражение (6), описывающее масштабирование параметров банка фильтров, принимает вид:

$$\omega_{step} = 2\pi/R,\ \omega_{bw} = \alpha\omega_{step}. \quad (11)$$

Параметры синусоидальной модели, необходимые для вычисления ФФКП, извлекаются из сигнала при помощи многоскоростной схемы, показанной на рисунке 2.

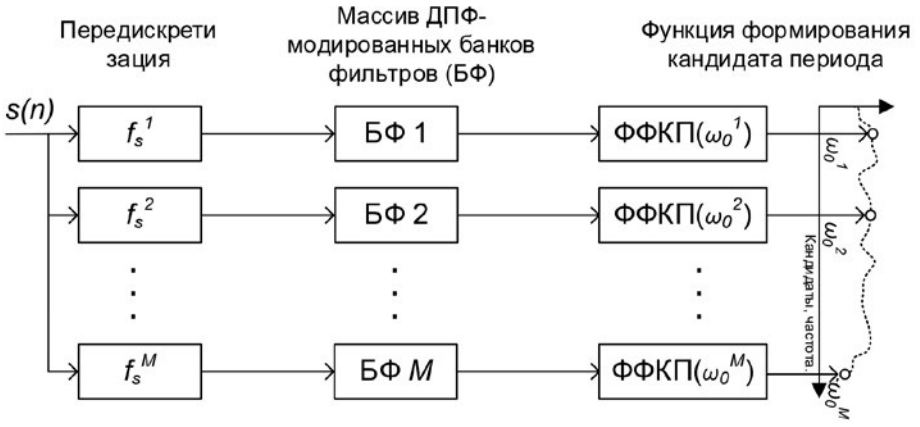

Рис. 2. Многоскоростная схема вычисления функции формирования кандидата периода основного тона (M — количество кандидатов периода)

2. ФУНКЦИЯ ФОРМИРОВАНИЯ КАНДИДАТА ПЕРИОДА ОСНОВНОГО ТОНА

В качестве функции формирования кандидата периода, как правило, используются различные метрики, основанные на автокорреляционной функции. Например, в [12] используется нормированная кросскорреляционная функция

$$\phi(n,l) = \frac{\sum_{i=0}^{N+l-1} s(n)s(n+l)}{\sqrt{e(0)e(l)}}, \quad (12)$$

где l — задержка в отсчетах, $e(l) = \sum_{i=0}^{N+l-1} s(n)^2$.

Функция (12) усредняет данные внутри кадра анализа и поэтому дает сглаженные значения. Чтобы улучшить временное разрешение, в [7] предложено использовать нормированную кросскорреляционную функцию на основе синусоидальной модели сигнала:

$$\phi_{inst}(n,l) = \frac{\sum_{k=1}^{K}[A_k(n)]^2 \cos(\omega_k(n)l)}{\sum_{k=1}^{K}[A_k(n)]^2}. \qquad (13)$$

В данной функции предполагается, что ширина полос фильтров анализа уже, чем минимально допустимое значение частоты основного тона, вследствие чего каждая гармоника сигнала всегда попадает в отдельный канал.

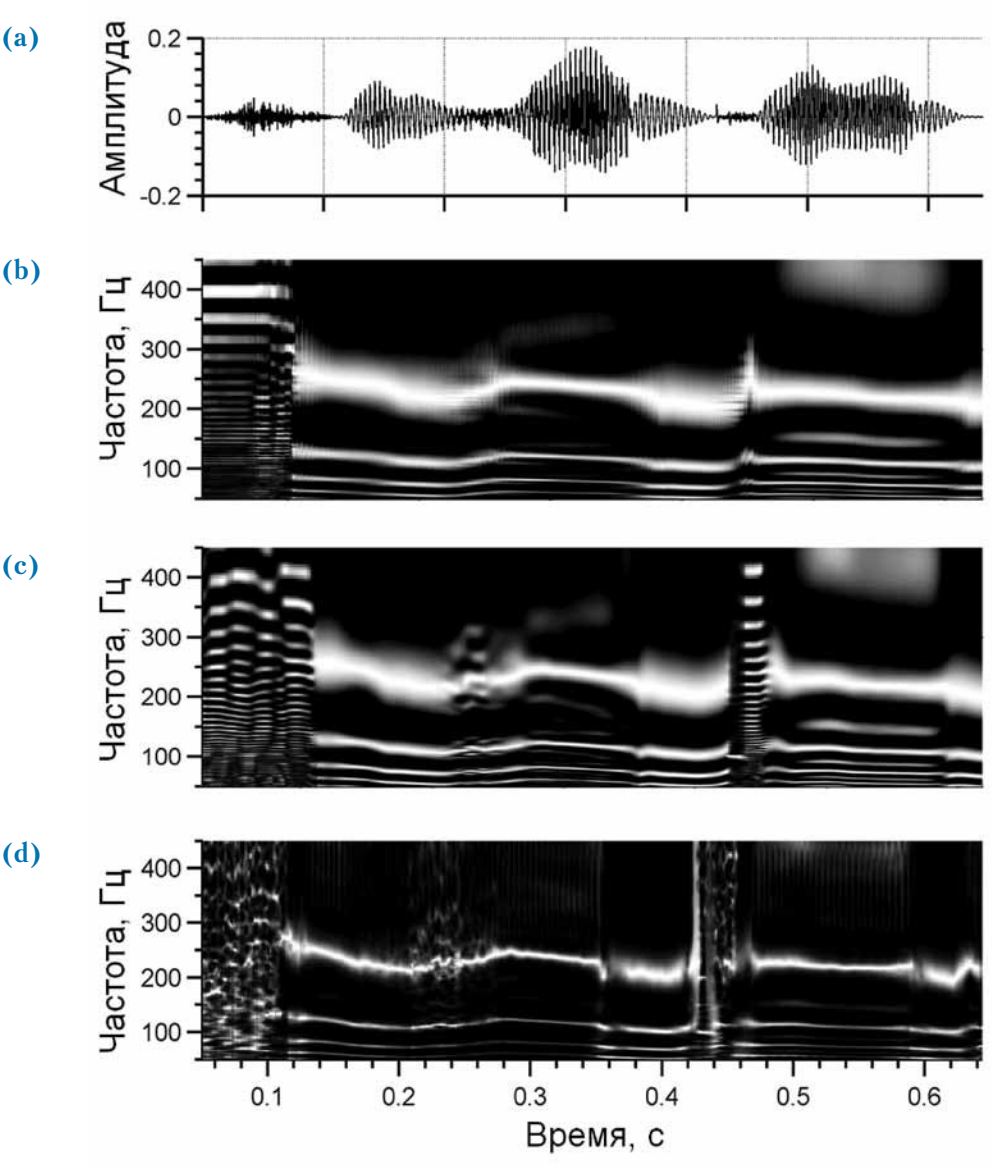

Рис. 3. Формирование кандидатов периода: (a) — исходный сигнал, (b) — ФФКП $\phi()$, (c) — ФФКП, полученная на основе синусоидальной модели [7] $\phi_{inst}()$, (d) — предлагаемая ФФКП $\phi_{ms}()(V=1)$

Многоскоростная схема анализа, описанная выше, подвержена явлению смешивания гармоник в каналах, которое возникает для высокочастотных кандидатов периода основного тона, когда обрабатываемый голос является низким. Вследствие этого появляются редкие единичные выбросы в высокочастотной области $\varphi_{inst}(n, l)$. Для того, чтобы уменьшить влияние эффекта смешивания гармоник, предлагается использовать следующую функцию формирования кандидата периода, которая использует мгновенные параметры, полученные для $2V + 1$ смежных отсчетов:

$$\phi_{ms}(n,l) = \prod_{\nu=-V}^{V} \sum_{k=1}^{K} A_k(n+\nu)\cos\left(\omega_k(n+\nu)l\right). \qquad (14)$$

В каждом отдельном канале схемы на рисунке 2 выполняется вычисление (14), отвечающее за конкретное значение задержки $l = 1/w_0^m$, где индекс $m = 1, \ldots M$, для чего в канал подается кадр сигнала, передискретизированный с коэффициентом R/l (в соответствии с (8)). Уравновешивание значений $\phi_{ms}(n, l)$ для различных кандидатов выполняется путем нормализации к единичной энергии каждого передискретизированного кадра сигнала. Использование в (14) амплитуд, не возведенных в квадрат, в отличие от (13) позволяет сделать вклад амплитуд различных гармоник более сбалансированным. Как правило, эффект смешивания гармоник возникает на коротких временных периодах и может быть существенно уменьшен умножением нескольких термов вида $\sum_{k=1}^{K} A_k(n)\cos(w_k l)$. На рисунке 3 показаны кандидаты периодов, сформированные функциями $\phi(n, l)$, $\phi_{inst}(n, l)$ и предлагаемой функцией для короткого речевого фрагмента. Очевидно, что функция $\phi_{ms}(n, l)$ имеет более высокое частотное и временное разрешение по сравнению с $\phi(n, l)$ и $\phi_{inst}(n, l)$.

3. АЛГОРИТМ ОЦЕНКИ ЧАСТОТЫ ОСНОВНОГО ТОНА

Предлагаемый алгоритм оценки частоты основного тона состоит из следующих шагов[1]:

1) выполнить передискретизацию фрейма входного сигнала $s(n)$ для каждого кандидата периода основного тона с частотой дискретизации (8);
2) нормировать энергию каждого передискретизированного фрейма к 1;
3) оценить мгновенные параметры синусоидальной модели согласно выражениям (2)–(5). Данный шаг повторяется для $2V + 1$ перекрывающихся кадров каждого фрейма. В качестве оконной функции в (2) использовать окно Хемминга;
4) вычислить функцию формирования кадидатов периода частоты основного тона (14), используя соответствующий набор параметров;
5) умножить полученное значение функции формирования кандидатов периода основного тона на взвешивающую функцию для ограничения низкочастотных кандидатов периода: $w_{wight}(\omega_0) = 0{,}2\frac{\omega_0}{\pi} + 0{,}8$;

[1] Условное название алгоритма «halcyon». Matlab-реализация алгоритма доступна по ссылке http://dsp.tut.su/halcyon.html

6) поиск наилучшего непрерывного контура частоты основного тона методом динамического программирования, максимизирующего сумму ФФКП на локальной последовательности кадров; в результате данного шага выбирается лучший кандидат $\omega_{0,best}(n)$, который является грубой (начальной) оценкой частоты основного тона;
7) вычислить уточненную оценку основного тона $\omega_{0,fine}(n)$, используя мгновенные параметры синусоидальной модели полученные для лучшего кандидата:

$$\omega_{0,fine}(n) = \frac{1}{\sum_{k=1}^{K} A_k(n)} \sum_{k=1}^{K} \frac{1}{k} \omega_k(n) A_k(n). \qquad (15)$$

Вычислительная сложность алгоритма (число умножений, требуемое для оценки одного значения частоты основного тона) является невысокой, поскольку в основе реализации банка фильтров лежит быстрое преобразование Фурье. Ориентировочно вы-числительная сложность оценивается как

$$O\,(IKN + 2(V+1)\,KNlog\,(N),$$

где I — порядок НЧ фильтра, используемого при децимации/интерполяции в процессе передискретизации.

Для практической реализации алгоритма были использованы следующие значения параметров: $K = 8$, $L = 4$, $R = 2K + 1 = 17$, $N = 68$, $M = 100$, $I = 121$, $V = 1$. Диапазон поиска частоты основного тона — 50–450 Гц. Данный диапазон разбит линейно в логарифмическом масштабе на 100 интервалов, каждый из которых соответствует одному кандидату периода частоты основного тона. Длительность перидескритезирован-ных кадров варьируется от 80 (самый низкочастотный кандидат) до 9 мс (самый высокочастотный кандидат).

4. ЭКСПЕРИМЕНТАЛЬНАЯ ОЦЕНКА ТОЧНОСТИ АЛГОРИТМА

Предлагаемый оценщик частоты основного тона (Halcyon) сравнивался с пятью известными и широко применяемыми алгоритмами RAPT [12], YIN [13], SWIPE' [14], IRAPT [7], PEFAC [15]. Сравнение выполнялось в терминах: 1) процента грубых ошибок (gross pitch error — GPE) и 2) среднего значения мелких ошибок (mean fine pitch error — MFPE). GPE вычисляется как процент вокализованных фреймов с ошибкой в оценке основного тона, превышающей от настоящего значения основного тона, при вычислении MFPE фреймы, содержащие грубые ошибки, не учитывались.

Для оценки временного разрешения алгоритма и его устойчивости к быстрому изменению основного тона были синтезированы модельные сигналы с изменяющейся частотой основного тона в диапазоне от 100 до 350 Гц. Полученные экспериментальные результаты были разделены на 6 групп в зависимости от скорости изменения частоты основного тона, измеряемой в процентах изменения тона на миллисекунду (0–0,3, 0,3–0,6, 0,6–0,9, 0,9–1,2, 1,2–1,5, >1,5). Усреднённые значения ошибок показаны на рисунке 4.

Алгоритмы IRAPT и Halcyon показывают более высокую устойчивость к изменению частоты основного тона — процент грубых ошибок для них остается незначительным до значения 1,5 %/ мс. График MFPE показывает, что предлагаемый алгоритм превосходит все остальные по частотно/временному разрешению. На рисунке 5 приведен пример анализа модельного сигнала с быстрым изменением тона. Алгорит-

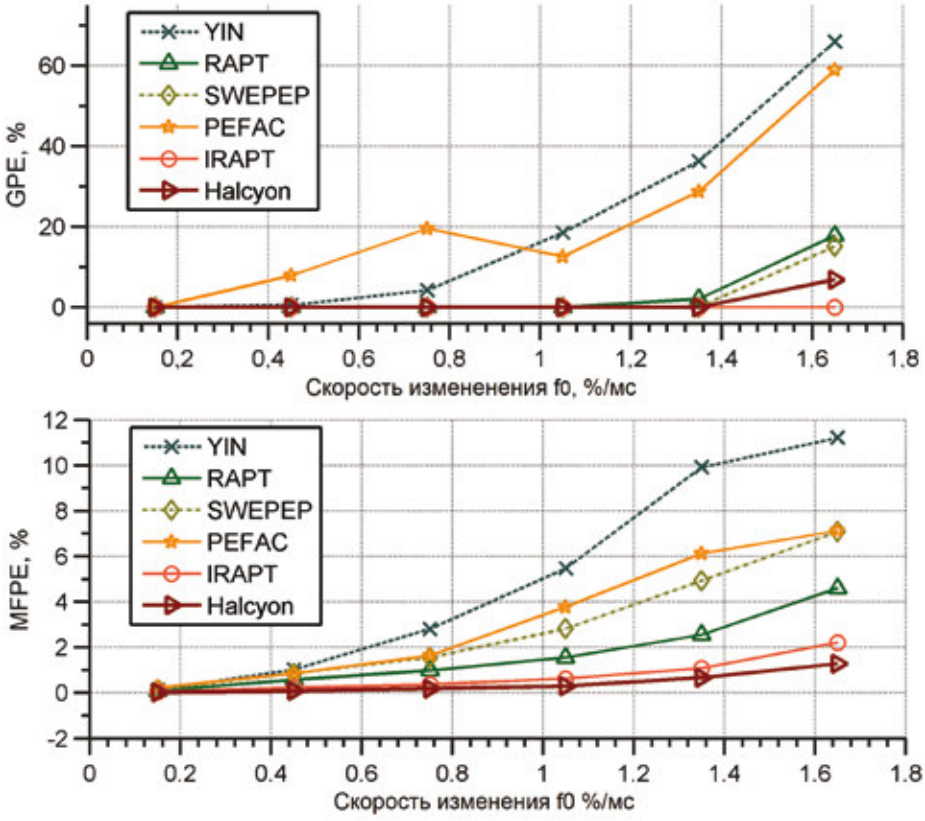

Рис. 4. Оценка временного разрешения алгоритмов выделения основного тона

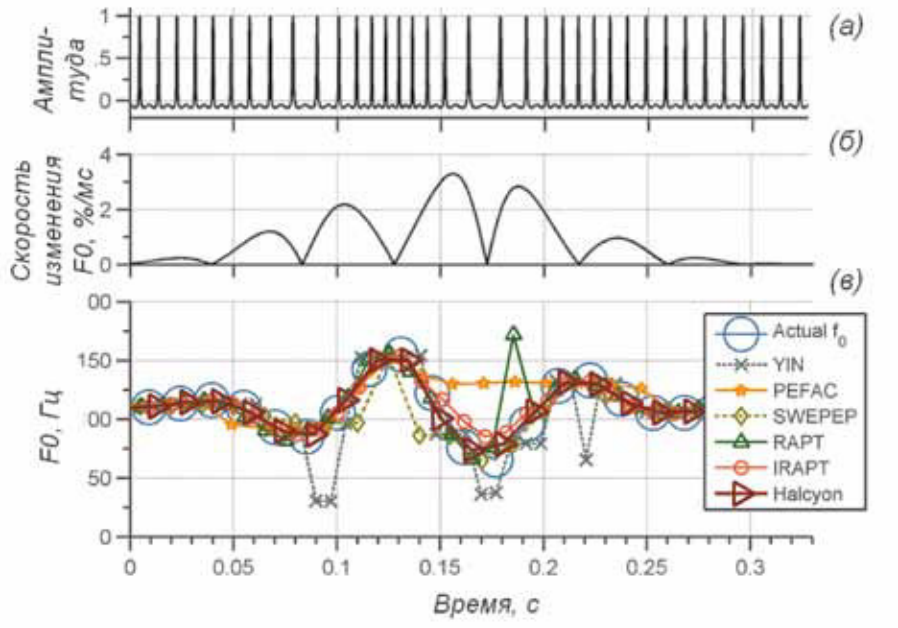

Рис. 5. Анализ сигнала с изменяемым тоном. (a) — исходный сигнал, (b) — скорость изменения основного тона, (c) — настоящий и рассчитанный контур частоты основного тона

мы IRAPT, SWIPE' и Halcyon дают оценку, весьма близкую к истинным значениям, остальные алгоритмы не демонстрируют такой точности.

Для экспериментов с использованием натуральной речи использовалась база PTDB-TUG [16]. Усредненные результаты экспериментов для чистой речи приведены в таблице 1. По сравнению с IRAPT 1 предлагаемый алгоритм демонстрирует в два раза меньший процент грубых ошибок (GPE) благодаря многоскоростной схеме анализа.

Таблица 1

Сравнение алгоритмов оценки частоты основного тона с использованием речевых сигналов

	Мужской голос		Женский голос	
	GPE%	MFPE%	GPE%	MFPE%
RAPT	3.687	1.737	6.068	1.184
YIN	3.184	1.389	3.960	0.835
SWIPE'	0.756	1.505	4.273	0.800
PEFAC	20.521	1.383	31.192	0.972
IRAPT 1	1.625	1.608	3.777	0.977
Halcyon	0.743	1.268	3.600	1.039

Для проверки на устойчивость к шумам к чистым речевым сигналам добавлялся шум двух видов (белый и речеподобный (англ. *babble*)) с различным значением ОСШ от –20 до 20 дБ. Усредненные результаты оценки тона в зашумленной речи показаны на рисунках 6,7.

Для белого шума все алгоритмы за исключением RAPT показывают хороший результат для ОСШ-10 дБ и выше. Для речеподобного шума результаты работы алгоритмов быстро ухудшаются, начиная со значения ОСШ 0 дБ. В целом предлагаемый алгоритм показывает приемлемую устойчивость к аддитивным шумам, учитывая, что в нем используются значения параметров синусоидальной модели, оцененные на очень коротких фреймах анализа (до 9 мс) для высокочастотных кандидатов частоты основного тона.

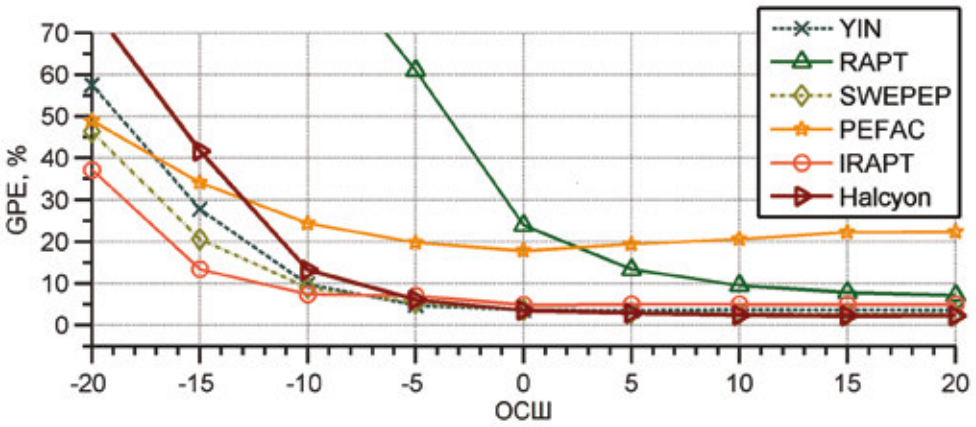

Рис. 6. Точность измерения основного тона (аддитивный белый шум)

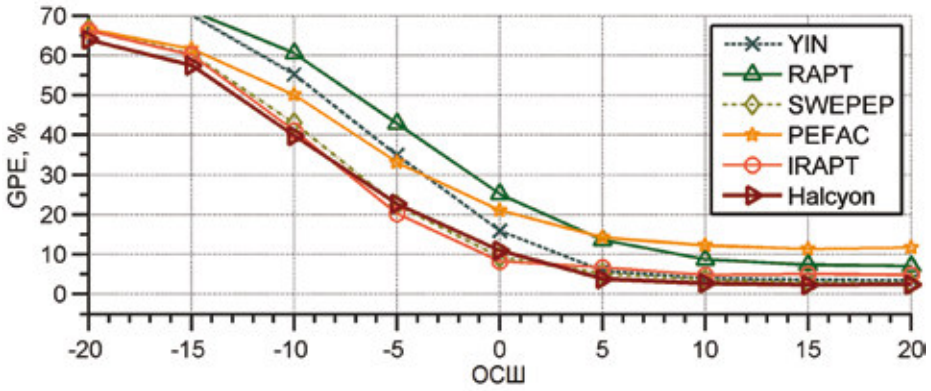

Рис. 7. Точность измерения основного тона (аддитивный речеподобный шум)

ЗАКЛЮЧЕНИЕ

В работе представлен алгоритм выделения частоты основного тона, который может быть применен в задачах анализа искусственно сгенерированных и речевых сигналов. В алгоритме выполняется декомпозиция сигнала на узкополосные составляющие, каждая из которых описывается синусоидальной моделью с мгновенными параметрами амплитуды, частоты и фазы. Для каждого кандидата периода частоты основного тона выполняется масштабирование анализирующего банк фильтров, что способствует точной оценке тона как для низких, так и для высоких голосов. Экспериментальные результаты показывают значительную устойчивость предлагаемого алгоритма к модуляциям основного тона, а также его высокое частотное и временное разрешение.

Благодарность

Работа выполнялась при поддержке компании ITForYou (Москва), а также Белорусского республиканского фонда фундаментальных исследований (грант № Ф17У-003).

ЛИТЕРАТУРА

1. *F. Zhang, G. Bi, Y. Q. Chen,* "Harmonic transform," in Vision, Image and Signal Processing, IEE Proceedings, vol.151, no.4, pp.257–263, 2004.
2. *R. J. McAulay, T. F. Quatieri,* "Speech analysis/synthesis based on a sinusoidal representation," IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 34, no. 4, pp. 744–754, 1986.
3. *J. Laroche, Y. Stylianou, E. Moulines,* "HNS: Speech modification based on a harmonic+noise model," in ICASSP-93 — IEEE International Conference on Acoustic, Speech, and Signal Processing, April 27-30, Minneapolis, USA, Proceedings, 1993. — pp. 550–553.
4. *J. O. Hong, P. J. Wolfe,* "Model-based estimation of instantaneous pitch in noisy speech," in INTERSPEECH 2009 — 10th Annual Conference of the International

Speech Communication Association, September 6–10, Brighton, UK, Proceedings, 2009. — Pp. 112–115.
5. *B. Resch, M. Nilsson, A. Ekman and W. B. Klejin* "Estimation of the Instanta-neous Pitch of Speech," IEEE Transactions on Audio, Speech and Language Processing, vol. 15, No. 15, Pp. 819–822, 2007.
6. *T. Abe, T. Kobayashi, S. Imai,* "Harmonics tracking and pitch extraction based on instantaneous frequency," in ICASSP-95 — IEEE International Conference on Acoustic, Speech, and Signal Processing, May 9-12, Detroit, USA, Proceedings, 1995. — Pp. 756–759.
7. *E. Azarov, M. Vashkevich, A. Petrovsky,* "Instantaneous pitch estimation based on RAPT framework," in EUSIPCO'12 — European Signal Processing Conference, August 27-31, Bucharest, Romania, Proceedings, 2012. — Pp. 2787–2791.
8. *E. Azarov, M. Vashkevich, A. Petrovsky,* "Instantaneous harmonic representa-tion of speech using multicomponent sinusoidal excitation," in INTERSPEECH 2013 –14th Annual Conference of the International Speech Communication Association, August 25–29, Lyon, France, Proceedings, 2013. — Pp. 1697–1701.
9. *E. Azarov, M. Vashkevich, A. Petrovsky,* "Guslar: A framework for automated singing voice correction," in ICASSP-2014 — IEEE International Conference on Acoustic, Speech, and Signal Processing, May 4–9, Florence, Italy, Proceedings, 2014. — Pp. 7919–7923.
10. K. Hotta, K. Funaki, "On a Robust F0 Estimation of Speech based on IRAPT using Robust TV-CAR Analysis," in APSIPA 2014 — Annual Summit and Conference Asia-Pacific Signal and Information Processing Association, 2014, December 9–12, Siem Reap, Cambodia, Proceedings, 2014. — Pp. 1–4.
11. *E. van den Berg, B. Ramabhadran,* "Dictionary-based pitch tracking with dy-namic programming," in INTERSPEECH 2014 –15th Annual Conference of the International Speech Communication Association, September 14–18, Singapore, Proceedings, 2014. — Pp. 1347–1351.
12. *D. Talkin,* "A Robust Algorithm for Pitch Tracking (RAPT)" in "Speech Cod-ing & Synthesis," W B Kleijn, K K Paliwal eds, Elsevier ISBN 0444821694, 1995.
13. *A. Cheveigné, H. Kawahara* "YIN, a fundamental frequency estimator for speech and music," Journal of the Acoustical Society of America, vol. 111, No. 4, Pp. 1917–1930, 2002.
14. *A. Camacho, J. G. Harris,* "A sawtooth waveform inspired pitch estimator for speech and music," Journal of the Acoustical Society of America, Vol. 123, No. 4, Pp. 1638–1652, 2008.
15. *S. Gonzalez, M. Brookes,* "PEFAC — A Pitch Estimation Algorithm Robust to High Levels of Noise," IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 22, No.2, Pp. 518–530, 2014.
16. *G. Pirker, M. Wohlmayr, S. Petrik, F. Pernkopf* "A Pitch Tracking Corpus with Evaluation on Multipitch Tracking Scenario," in INTERSPEECH 2011 — 12th Annual Conference of the International Speech Communication Association, August 28–31, Lyon, France, Proceedings, 2011. — Pp. 1509–1512.

ESTIMATION OF INSTANTANEOUS FUNDAMENTAL FREQUENCY OF SPEECH BASED ON MULTIRATE SIGNAL PROCESSING

Maksim I. Vashkevich,
Candidate of technical Sciences, associate Professor of the Belarusian state University of Informatics and Radioelectronics (BSUIR)

Iliy S. Azarov,
Doctor of technical Sciences, associate Professor, BSUIR

Aleksandr A. Petrovsky,
Doctor of technical Sciences, Professor of the chair of electronic computing BSUIR

Abstract

The paper presents an algorithm for accurate pitch estimation that takes advantage of the sinusoidal model with instantaneous parameters. The algorithm decomposes the signal into subband components, extracts their instantaneous parameters and evaluates period candidate generating function (PCGF). In order to achieve high accuracy for low and high-pitched sounds it is assumed that possible pitch variation range is proportional to current pitch value. The bandwidths of the decomposition filters and length of the analysis frame are scaled for each period candidate by multirate sampling. The algorithm is compared to other widely used pitch extractors on artificial quasiperiodic signals and natural speech. The proposed algorithm shows a remarkable frequency and time resolution for pitch-modulated sounds and performs well both in clean and noisy conditions.

Keywords: fundamental frequency, pitch, multirate signal processing