

ПРОГРЕССИВНИ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

ПРОГРЕССИВНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

PROGRESSIV INFORMATICS TECHNOLOGIES

УДК 004.934

Имамвердиев Я. Н.¹, Сухостат Л. В.²

¹Канд. техн. наук, зав. отделом, Институт информационных технологий Национальной Академии Наук Азербайджана, Баку, Азербайджан

²Научный сотрудник, Институт информационных технологий Национальной Академии Наук Азербайджана, Баку, Азербайджан

МЕТОД ОЦЕНКИ ПЕРИОДА ОСНОВНОГО ТОНА С ПРИМЕНЕНИЕМ ЭМПИРИЧЕСКОГО ВЕЙВЛЕТ ПРЕОБРАЗОВАНИЯ

Оценка периода основного тона речевого сигнала используется во многих важных областях применения речевых технологий. Однако среди существующих методов только некоторые могут работать для случая нелинейных и нестационарных сигналов. Основная причина состоит в том, что методы обнаружения периода основного тона основаны на предположении, что сам процесс речеобразования линеен, а речевые сигналы являются локально стационарными. Выбор алгоритма для оценки периода основного тона всегда состоит в нахождении компромисса между временем и разрешающей способностью по частоте, робастностью, задержкой и вычислительной сложностью. Целью данной работы является разработка метода оценки периода основного тона на основе эмпирического вейвлет преобразования. Метод построения семейства адаптированных вейвлетов состоит в предположении, что фильтры зависят от расположения информации в спектре анализируемого сигнала. Эмпирические вейвлеты определяются как полосовые фильтры на каждом отрезке речевого сигнала. Для определения периода основного тона в качестве признаков рассматривается мгновенная частота. Оператор разделения энергии Тигера-Кайзера используется для ее выделения. В работе приводится сравнение данного метода с другими алгоритмами оценки периода основного тона.

Ключевые слова: период основного тона, эмпирическое вейвлет преобразование, оператор разделения энергии Тигера-Кайзера, внутренняя модовая функция, мгновенная частота.

НОМЕНКЛАТУРА

AUTOC – Modified Autocorrelation Method;
MBSC – Multi-Band Summary Correlogram;
AMDF – Average magnitude difference function;
PEFAC – Pitch estimation Filter with Amplitude Compression;
IMF – Intrinsic Mode Function;
HHT – Hilbert-Huang Transform;
EMD – Empirical Mode Decomposition;
EWT – Empirical Wavelet Transform;
ТКЕО – Teager-Kaiser Energy Operator;
GPE – Gross Pitch Error;
SNR – signal-to-noise ratio;
MFPE – Mean Fine Pitch Error;
DESA – Discrete Energy Separation Algorithm;
DESA-1 – первый дискретный алгоритм разделения энергии;
DESA-2 – второй дискретный алгоритм разделения энергии;
AM-FM – Amplitude-Modulation Frequency-Modulation;

IF1 – мгновенная частота из первой IMF;
 T_n – переходная фаза;
 N_{FPE} – число вокализованных участков без GPE;
 ω_n – границы между сегментами;
 f_k – функция IMF;
 $F0_{true}(i)$ – эталонное значение периода основного тона;
 $F0_{est}(i)$ – полученное значение периода основного тона;
 i – число фреймов;
 N_{VU} – число фреймов, в которых полученное значение основного тона и эталонное значение определены как вокализованные;
 $\alpha_j(t)$ – мгновенная амплитуда;
 $\theta_j(t)$ – фазовая функция;
 $H[c_j(t)]$ – преобразование Гильберта;
 $z_j(t)$ – аналитический сигнал;
 $X(t)$ – исходный сигнал;
 $\text{Re}\{\}$ – реальная часть комплексного выражения;

$H(\omega, t)$ – спектр Гильберта;

$\psi(\cdot)$ – оператор ТКЕО;

$x(t)$ – сигнал;

$\dot{x}(t)$ – производные первого порядка;

$\ddot{x}(t)$ – производные второго порядка;

$f(n)$ – мгновенная частота;

$|\alpha(n)|$ – мгновенная амплитуда;

F_0 – период основного тона;

N – число сегментов;

Λ_n – сегмент сигнала;

$C^k([0,1])$ – пространство k раз дифференцируемых функций на интервале $[0,1]$;

$\beta(x)$ – функция из $C^k([0,1])$;

γ – параметр определяющий размер границ;

$\hat{f}_n(\omega)$ – эмпирическая масштабируемая функция;

$\hat{\psi}_n(\omega)$ – эмпирические вейвлеты;

$W_{f_j}^{\varepsilon}(0, t)$ – детализирующие коэффициенты;

$W_{f_j}^{\varepsilon}(0, t)$ – аппроксимирующие коэффициенты.

ВВЕДЕНИЕ

Начиная с 1970-х годов было разработано множество алгоритмов для выделения периода основного тона. Среди них можно выделить модифицированный автокорреляционный метод (AUTOC) [1], кепстральный метод, многополосную агрегацию коррелограмм (MBSC) [2], BaNa [3], YIN [4], YAAPT [5], среднее значение разностной функции (AMDF) [1], SWIPE' [6] и метод оценки основного тона на основе амплитудного сжатия (PEFAC) [7].

Методы оценки периода основного тона можно, в общем, разделить на временные, частотные и гибридные. Первая категория, в основном, ищет пики в автокорреляционной функции, вторая – в спектре мощности, тогда как третья группа выполняет частотно-временной анализ выхода банка полосовых фильтров. Во многих случаях алгоритмы выделяют кандидатов периода основного тона для каждого временного фрейма, а затем используют временные ограничения на непрерывность.

Однако только некоторые из них могут работать для случая нелинейных и нестационарных сигналов.

Некоторые хорошие альтернативы – методы на основе оценки параметров мгновенной частоты. Мгновенная частота была представлена в [8, 9], оценка методов предложена в [8–13]. Она необходима для понимания подробных механизмов обработки нелинейных и нестационарных процессов. На практике мгновенная частота обычно вычисляется из внутренней модовой функции (IMF) с помощью преобразования Гильберта-Хунга (ННТ). ННТ для получения IMF применяет эмпирическую модовую декомпозицию (EMD) [13]. В сравнении с методом автокорреляции данный подход более точен и быстрее обнаруживает период основного тона. Алгоритм EMD, применяемый при вычислении IMF, адаптивен, его вычислительная сложность, включая число извлеченных IMF и количество вычислений на этапе их получения, строго зависит от сложности самих речевых сигналов. Но метод EMD автоматически оценивает число мод, что существенно влияет на вычислительную

сложность. В связи с этим в работе предлагается метод эмпирического вейвлет преобразования (EWT) [14].

Далее для получения мгновенной частоты применяется преобразование Гильберта. Однако оператор разделения энергии Тигера-Кайзера (ТКЕО) [15] превосходит его по вычислительной сложности и скорости на реальных сигналах. Преобразование Гильберта и оператор ТКЕО могут быть применены только к монокомпонентным сигналам. В случае мультикомпонентных сигналов необходимо разбиение сигнала на простые компоненты перед применением методов. Для этого удобно использовать узкополосные фильтры [10]. Однако, в случае частотно-модулированных компонентов, это не всегда возможно в силу широкого диапазона частот.

Целью данной работы является разработка нового метода оценки периода основного тона на основе мгновенной частоты с применением EWT и нелинейного оператора ТКЕО. Приводятся результаты практических экспериментов на речевой базе данных Keele [16]. Оценивается робастность предложенного метода к шуму.

1 ПОСТАНОВКА ЗАДАЧИ

Пусть рассматривается метод построения семейства вейвлетов адаптированных к обрабатываемым сигналам. Представлены реальные сигналы, где спектр симметричен относительно частоты $\omega = 0$, а также нормализованная ось Фурье, которая имеет периодичность для того, чтобы удовлетворить критерию Шеннона, и ограничивается отрезком $\omega \in [0, \pi]$.

Отрезок $[0, \pi]$ делится на N смежных сегментов. Каждый сегмент $\Lambda_n = [\omega_{n-1}, \omega_n]$ (где $\omega_0 = 0$ и $\omega_N = \pi$). Вокруг ω_n определяется переходная фаза T_n шириной $2\tau_n$.

Тогда эмпирические вейвлеты [14] определяются как полосовые фильтры на каждом Λ_n . Банки фильтров стро-

ятся согласно условию $\gamma < \min_n \left(\frac{\omega_{n+1} - \omega_n}{\omega_{n+1} + \omega_n} \right)$. Далее оп-

ределяются компоненты IMF.

После получения IMF компонент для вычисления мгновенной частоты и мгновенной амплитуды к каждой IMF применяется преобразование Гильберта [17] (дано: реальный сигнал с симметричным спектром относительно $\omega = 0$, надо: оценка периода основного тона на основе мгновенной частоты).

2 ЛИТЕРАТУРНЫЙ ОБЗОР

Вейвлет-анализ в настоящее время является одним из средств исследования данных. Он позволяет решать задачи моделирования, идентификации, аппроксимировать стационарные и нестационарные процессы. Преимущества вейвлетов включают частотно-временную локализацию сигналов.

Нестационарные и нелинейные сигналы требуют адаптивные пути решения. Адаптивность состоит в рассмотрении тех областей сигнала, где содержится полезная информация. Вейвлет преобразования способны собирать основную информацию о сигнале в небольшое число вейвлет коэффициентов. Полезное свойство вейвлет преобразования состоит и в том, что его можно рассматривать как применение банка фильтров.

Известен подход, называемый вейвлетами Малвара-Вильсона [18, 19], который пытается построить адаптивное представление путем сегментации сигнала во временной области, чтобы выделить временные интервалы, содержащие спектральную информацию о сигнале. В то время как данный подход интересен, временная сегментация является трудно реализуемой задачей.

В [20] авторы предлагают метод, называемый «brushlets», который строит адаптивный банк фильтров в области Фурье. Он использует идею вейвлетов Малвара-Вильсона, но сегментирует спектр Фурье вместо самого сигнала. Данная задача является конструктивно довольно сложной.

В недавней работе И. Добеши под названием «synchrosqueezed всплески» [21] предлагается подход, сочетающий в себе классический вейвлет-анализ и метод перераспределения частотно-временной информации. Этот алгоритм позволяет получить более точное частотно-временное представление.

В основу каждого метода положено классическое вейвлет преобразование, но каждый из них имеет свои индивидуальные особенности.

3 МАТЕРИАЛЫ И МЕТОДЫ

Эмпирическая масштабируемая функция и эмпирические вейвлеты определяются следующим образом

$$\hat{\phi}_n(\omega) = \begin{cases} 1, & \text{если } |\omega| \leq \omega_n - \tau_n; \\ \cos \left[\frac{\pi}{2} \beta \left(\frac{1}{2\tau_n} (|\omega| - \omega_n + \tau_n) \right) \right], & \text{если } \omega_n - \tau_n \leq |\omega| \leq \omega_n + \tau_n; \\ 0, & \text{в противном случае} \end{cases} \quad (1)$$

и

$$\hat{\psi}_n(\omega) = \begin{cases} 1, & \text{если } \omega_n + \tau_n \leq |\omega| \leq \omega_{n+1} - \tau_{n+1}; \\ \cos \left[\frac{\pi}{2} \beta \left(\frac{1}{2\tau_{n+1}} (|\omega| - \omega_{n+1} + \tau_{n+1}) \right) \right], & \text{если } \omega_{n+1} - \tau_{n+1} \leq |\omega| \leq \omega_{n+1} + \tau_{n+1}; \\ \sin \left[\frac{\pi}{2} \beta \left(\frac{1}{2\tau_n} (|\omega| - \omega_n + \tau_n) \right) \right], & \text{если } \omega_n - \tau_n \leq |\omega| \leq \omega_n + \tau_n; \\ 0, & \text{в противном случае.} \end{cases} \quad (2)$$

Функция $\beta(x)$, произвольная из $C^k([0,1])$ – функция такая, что

$$\beta(x) = \begin{cases} 0, & \text{если } x \leq 0 \text{ и } \beta(x) + \beta(1-x) = 1, \forall x \in [0,1]; \\ 1, & \text{если } x \geq 1, \end{cases}$$

τ_n выбирается пропорционально ω_n : $\tau_n = \gamma \omega_n$, где $0 < \gamma < 1$. Следовательно, для всех $n > 0$ уравнения (1) и (2) принимают вид

$$\hat{\phi}_n(\omega) = \begin{cases} 1, & \text{если } |\omega| \leq (1-\gamma)\omega_n; \\ \cos \left[\frac{\pi}{2} \beta \left(\frac{1}{2\gamma\omega_n} (|\omega| - (1-\gamma)\omega_n) \right) \right], & \text{если } (1-\gamma)\omega_n \leq |\omega| \leq (1+\gamma)\omega_n; \\ 0, & \text{в противном случае} \end{cases} \quad (3)$$

и

$$\hat{\psi}_n(\omega) = \begin{cases} 1, & \text{если } (1+\gamma)\omega_n \leq |\omega| \leq (1-\gamma)\omega_{n+1}; \\ \cos \left[\frac{\pi}{2} \beta \left(\frac{1}{2\gamma\omega_{n+1}} (|\omega| - (1-\gamma)\omega_{n+1}) \right) \right], & \text{если } (1-\gamma)\omega_{n+1} \leq |\omega| \leq (1+\gamma)\omega_{n+1}; \\ \sin \left[\frac{\pi}{2} \beta \left(\frac{1}{2\gamma\omega_n} (|\omega| - (1-\gamma)\omega_n) \right) \right], & \text{если } (1-\gamma)\omega_n \leq |\omega| \leq (1+\gamma)\omega_n; \\ 0, & \text{в противном случае.} \end{cases} \quad (4)$$

Теперь можем определить эмпирическое вейвлет преобразование $W_f^E(n,t)$ также как и в случае классического вейвлет преобразования:

$$W_f^E(n,t) = \langle f, \psi_n \rangle = \int f(\tau) \overline{\psi_n(\tau-t)} d\tau = \hat{f}(\omega) * \overline{\hat{\psi}_n(\omega)},$$

а аппроксимирующие коэффициенты – скалярных произведений с масштабируемой функцией

$$W_f^E(0,t) = \langle f, \phi_1 \rangle = \int f(\tau) \overline{\phi_1(\tau-t)} d\tau = \hat{f}(\omega) * \overline{\hat{\phi}_1(\omega)},$$

где $\hat{\psi}_n(\omega)$ и $\hat{\phi}_1(\omega)$ определяются из уравнений (3) и (4) соответственно. Обратное преобразование принимает вид

$$\begin{aligned} f(t) &= W_f^E(0,t) * \phi_1(t) + \sum_{n=1}^N W_f^E(n,t) * \psi_n(t) = \\ &= \hat{W}_f^E(0,\omega) * \hat{\phi}_1(\omega) + \sum_{n=1}^N \hat{W}_f^E(n,\omega) * \hat{\psi}_n(\omega). \end{aligned}$$

Функция IMF f_k определяется следующим образом:

$$f_0(t) = W_f^E(0,t) * \phi_1(t),$$

$$f_k(t) = W_f^E(k,t) * \psi_k(t).$$

Использование преобразования Гильберта позволяет получить для каждого момента времени

$$H[c_j(t)] = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{c_j(\tau)}{t-\tau} d\tau.$$

Теперь можно построить аналитический сигнал $z_j(t)$ как

$$z_j(t) = c_j(t) + i H[c_j(t)],$$

который может быть представлен в виде

$$z_j(t) = \alpha_j(t) \exp(i\omega_j(t)).$$

Мгновенная амплитуда $\alpha_j(t)$ и фазовая функция $\theta_j(t)$ определяются как

$$\alpha_j(t) = \sqrt{c_j^2(t) + H^2[c_j(t)]},$$

$$\theta_j(t) = \arctan \frac{H[c_j(t)]}{c_j(t)}.$$

Кроме того, мгновенная частота $\omega_j(t)$ может быть вычислена следующим образом:

$$\omega_j(t) = \frac{d\theta_j(t)}{dt}.$$

Так исходный сигнал может быть представлен в форме (5):

$$X(t) = \operatorname{Re} \sum_{j=1}^n \alpha_j(t) \exp[i\omega_j(t)dt], \quad (5)$$

где остаток опущен.

Выражение (5) позволяет представить мгновенную амплитуду и частоту в трехмерном пространстве, где амплитуда – это высота в частотно-временной плоскости. Это частотно-временное распределение представлено как спектр Гильберта $H(\omega, t)$

$$H(\omega, t) = \operatorname{Re} \sum_{j=1}^n \alpha_j(t) \exp[i\omega_j(t)dt].$$

ННТ удовлетворяет требованию адаптивности для анализа нестационарных сигналов. Таким образом, сигнал может быть локально и точно отображен во временной частотной области путем применения спектра Гильберта.

Оператор разделения энергии Тигера-Кайзера – нелинейный оператор, который успешно применяется во многих инженерных приложениях [15]. Он обнаруживает модуляцию энергии и определяет мгновенную частоту и мгновенную амплитуду от АМ-ФМ сигнала [9]. Оператор ТКЕО $\Psi(\cdot)$ для сигнала $x(t)$ определяется как

$$\Psi[x(t)] = [\dot{x}(t)]^2 - x(t)\ddot{x}(t).$$

Оператор разделения энергии Тигера-Кайзера – нелинейный оператор, который успешно применяется во многих инженерных приложениях [15]. Он обнаруживает модуляцию энергии и определяет мгновенную частоту и мгновенную амплитуду от АМ-ФМ сигнала [9]. Оператор ТКЕО $\Psi(\cdot)$ для сигнала $x(t)$ определяется как

$$\Psi[x(t)] = [\dot{x}(t)]^2 - x(t)\ddot{x}(t).$$

В дискретно-временной области оператор принимает вид

$$\Psi[x(n)] = x^2(n) - x(n+1) \cdot x(n-1).$$

Мгновенная частота $f(n)$ и мгновенная амплитуда $|\alpha(n)|$ в любой момент времени для сигнала $x(n)$ даются как

$$y(n) = x(n) - x(n-1), \quad (6)$$

$$f(n) = \arccos \left(1 - \frac{\Psi[y(n)] + \Psi[y(n+1)]}{4\Psi[x(n)]} \right), \quad (7)$$

$$|\alpha(n)| = \sqrt{\frac{\Psi[x(n)]}{\sin^2[f(n)]}}, \quad (8)$$

$$f(n) = \frac{1}{2} \arccos \left(1 - \frac{\Psi[x(n+1)] - \Psi[x(n-1)]}{2\Psi[x(n)]} \right), \quad (9)$$

$$|\alpha(n)| = \frac{2\Psi[x(n)]}{\sqrt{\Psi[x(n+1)] - \Psi[x(n-1)]}}. \quad (10)$$

В общем, метод демодуляции (6)–(8) известен как DESA-1, а метод (9) и (10) – как DESA-2. Алгоритм DESA-2 только требует три значения для вычисления энергии в каждый момент времени и более прост в вычислении. Поэтому в данной работе мы рассматриваем алгоритм DESA-2.

4 ЭКСПЕРИМЕНТЫ

Для проведения экспериментов была рассмотрена речевая база данных Keele [16]. Речевые образцы получены от 10 дикторов (5 мужчин и 5 женщин). Также содержатся записи, извлеченные из ларингографа (laryngograph) с помощью алгоритма автокорреляции и сопровождаются измерениями F_0 .

Для тестирования устойчивости к внешним шумам рассматриваемых алгоритмов определения периода основного тона к сигналам добавляется белый шум с различными уровнями SNR. Для генерации зашумленной речи с определенным значением SNR, энергия сигнала вычисляется только на вокализованных участках речевого сигнала, и шум усиливается или ослабевает до определенного уровня, чтобы удовлетворить значению целевого SNR.

Период основного тона был оценен с помощью автокорреляции с окном 26,5 мс и сдвигом в 10 мс.

Для сравнения производительности методов обнаружения периода основного тона, используются следующие метрики ошибок [17]:

а) процент грубых ошибок (GPE) определяет соотношение кадров, рассчитанное на основе вокализованных участков полученного периода основного тона и эталонных значений (ground truth), для которых относительная погрешность оценки выше, чем определенный порог δ (обычно 20% [22]):

$$GPE = \frac{N_{F0E}}{N_{VT}} \times 100\%,$$

где N_{F0E} – число фреймов, для которых выполняется условие

$$\left| \frac{F0_{est}(i)}{F0_{true}(i)} - 1 \right| > \delta\%;$$

б) средний процент мелких ошибок (MFPE) вычисляется на вокализованных участках, где не наблюдаются ошибки GPE [23]:

$$MFPE = \frac{1}{N_{FPE}} \sum_{i=1}^{N_{FPE}} \frac{|F0_{true}(i) - F0_{est}(i)|}{F0_{true}(i)} \times 100\%,$$

где N_{FPE} – число вокализованных участков без GPE.

EWT разлагает сегменты сигнала на серию IMF функций для дальнейшего извлечения мгновенной частоты (рис. 1). Из всего набора IMF выбираем ту, чей период ближе к исходному сигналу. Из рис. 1 видно, что IMF1 содержит информацию о периоде основного тона и хорошо отображает форму сигнала. Блок-схема процесса на основе EWT-ТКЕО показана на рис. 2 (здесь IF1 обозначает мгновенную частоту получаемую из первой IMF).

5 РЕЗУЛЬТАТЫ

Эксперименты проводились в среде Matlab 2011b [24]. Сравнение производительности предложенного метода и других популярных алгоритмов обнаружения периода основного тона приводится в таблице 1.

Как показано в таблице 1, GPE для предложенного метода на основе EWT по сравнению с другими методами извлечения периода основного тона значительно меньше, чем у методов ННТ, YIN и SWIPE' при различных уровнях шума. Предложенный метод является более робастным по сравнению с другими методами и лучше работает при высоких SNR.

6 ОБСУЖДЕНИЕ

Предложенный метод оценки периода основного тона на основе эмпирического вейвлет преобразования обеспечивает существенную точность выделения высоты. Однако данный подход требует больших временных затрат.

По сравнению с другими алгоритмами оценки высоты, такими как YAAPT, ННТ и SWIPE', предложенный метод более точен. Метод был протестирован на устойчивость к внешним шумам.

Эффективность предложенного подхода может быть повышена при увеличении числа признаков. Однако это значительно влияет на скорость работы алгоритма.



Рисунок 2 – Схема оценки периода основного тона

Таблица 1 – Сравнение производительности методов извлечения периода основного тона

Метод		SNR (dB)			
		-5	0	10	15
SWIPE'	GPE	20,82	14,28	11,92	8,43
	MFPE	1,77	1,21	0,90	0,83
YAAPT	GPE	13,90	23,35	9,98	6,26
	MFPE	2,10	1,94	1,51	0,81
ННТ	GPE	19,30	6,02	11,34	5,10
	MFPE	0,56	0,43	0,35	0,33
EWT-ТКЕО	GPE	15,11	5,77	10,56	4,79
	MFPE	0,37	0,30	0,28	0,14

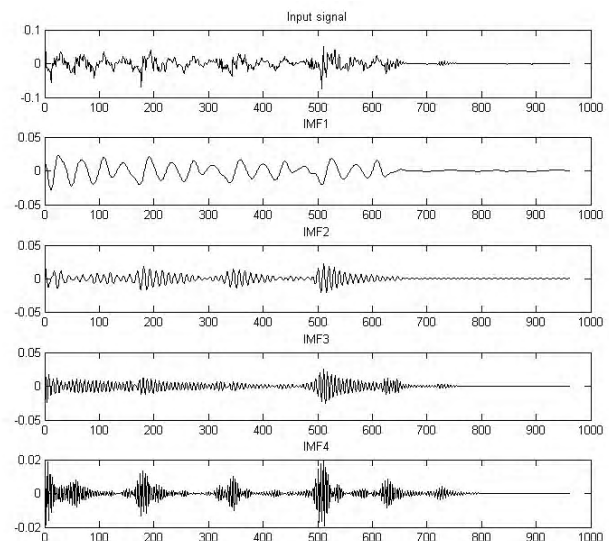
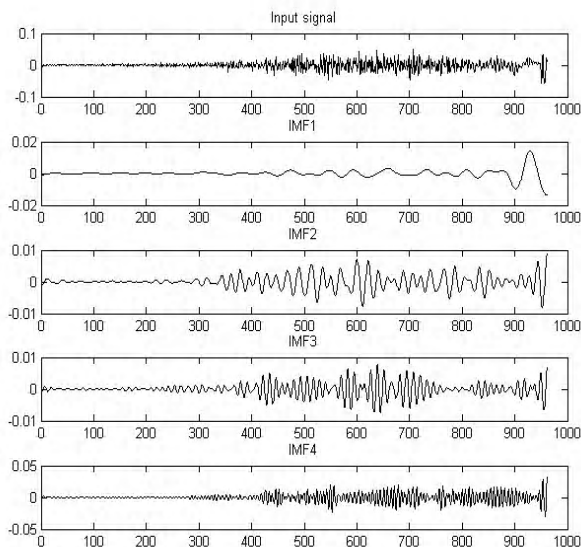


Рисунок 1 – Декомпозиция сигналов с помощью эмпирического вейвлет преобразования

ВЫВОДЫ

Решена актуальная задача разработки метода выделения периода основного тона речевого сигнала. Идея предлагаемого подхода состоит в том, что мгновенная частота содержит информацию о периоде основного тона. Был рассмотрен метод EWT. Для выделения мгновенной частоты был предложен оператор ТКЕО. Для проведения экспериментов белый шум был добавлен к речевому сигналу перед применением метода извлечения периода основного тона. Алгоритм протестирован при различных уровнях шума. Было показано, что точность алгоритма выше, чем у алгоритмов YAAPT, NHT и SWIPE' в случае быстрых модуляций высоты. Эксперименты показывают достаточную эффективность предлагаемого подхода при аддитивном шуме.

БЛАГОДАРНОСТИ

Данная работа выполнена при финансовой поддержке Фонда Развития Науки при Президенте Азербайджанской Республики – Грант № ЕЭФ-РЭТН-МҚМ-2/ЭКТ-2-2013-7(13)-29/18/1.

СПИСОК ЛИТЕРАТУРЫ

1. A comparative performance study of several pitch detection algorithms / [L. Rabiner, M. J. Cheng, A. E. Rosenberg, C. A. McGonegal] // *IEEE Transactions on Acoustics, Speech and Signal Processing*. – 1976. – № 5. – P. 399–417. DOI: 10.1109/TASSP.1976.1162846
2. Tan L. N. Multi-band summary correlogram-based pitch detection for noisy speech / L. N. Tan, A. Alwan // *Speech Communication*. – 2013. – Vol. 55, № 78. – P. 841–856. DOI: 10.1016/j.specom.2013.03.001
3. BaNa: a hybrid approach for noise resilient pitch detection / [H. Ba, N. Yang, I. Demirkol, W. Heinzelman] // *IEEE Statistical Signal Processing Workshop*. – 2012. – P. 369–372. DOI: 10.1109/SSP.2012.6319706
4. De Cheveigne A. Yin, a fundamental frequency estimator for speech and music / A. De Cheveigne, H. Kawahara // *Journal of the Acoustical Society of America*. – 2002. – Vol. 111, № 4. – P. 1917–1930. DOI: 10.1121/1.1458024
5. Kasi K. Yet another algorithm for pitch tracking / K. Kasi, S. A. Zahorian // *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*. – 2002. – Vol. 1. – P. 361–364. DOI: 10.1109/ICASSP.2002.5743729
6. Camacho A. SWIPE: a sawtooth waveform inspired pitch estimator for speech and music: thesis ... doctor of philosophy / Camacho A. – Florida: University of Florida, 2007.
7. Gonzalez S. PEFAC – A Pitch Estimation Algorithm Robust to High Levels of Noise / S. Gonzalez, M. Brookes // *IEEE Transactions on Audio, Speech and Language Processing*. – 2011. – Vol. 22, № 2. – P. 518–530. DOI: 10.1109/TASLP.2013.2295918
8. Boashash B. Estimating and interpreting the instantaneous frequency of a signal / B. Boashash // *Proceedings of the IEEE*. – 1992. – Vol. 80, № 4. – P. 520–538. DOI: 10.1109/5.135376
9. Maragos P. On amplitude and frequency demodulation using energy operators / P. Maragos, J. F. Kaiser, T. F. Quatieri // *IEEE Transactions on Signal Processing*. – 1993. – Vol. 41, № 4. – P. 1532–1550. DOI: 10.1109/78.212729
10. Abe T. Harmonics tracking and pitch extraction based on instantaneous frequency / T. Abe, T. Kobayashi, S. Imai // *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*. – 1995. – Vol. 1. – P. 756–759. DOI: 10.1109/ICASSP.1995.479804
11. Abe T. Sinusoidal model based on instantaneous frequency attractors / T. Abe, M. Honda // *IEEE Transactions on Audio, Speech and Language Processing*. – 2006. Vol. 14, № 4. – P. 1292–1300. DOI: 10.1109/TSA.2005.858545
12. Azarov E. Estimation of the instantaneous harmonic parameters of speech / E. Azarov, A. Petrovsky, M. Parfieniuk // *Proceedings of the European Signal Processing Conference*. – 2008.
13. The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis / [N. E. Huang, Z. Shen, S. R. Long and other] // *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*. – 1998. – Vol. 454, № 1971. – P. 903–995. DOI: 10.1098/rspa.1998.0193
14. Gilles J. Empirical Wavelet Transform / J. Gilles // *IEEE Transactions on Signal Processing*. – 2013. – Vol. 61, № 16. – P. 3999–4010. DOI: 10.1109/TSP.2013.2265222
15. Vakman D. On the analytic signal, the Teager–Kaiser energy algorithm, and other methods for defining amplitude and frequency / D. Vakman // *IEEE Transactions on Signal Processing*. – 1996. – Vol. 44, № 4. – P. 791–797. DOI: 10.1109/78.492532
16. Chu W. Reducing f0 frame error of f0 tracking algorithms under noisy conditions with an unvoiced/voiced classification frontend / W. Chu, A. Alwan // *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*. – 2009. – P. 3969–3972. DOI: 10.1109/icassp.2009.4960497
17. Varga A. Assessment for automatic speech recognition: II. Noisex-92: a database and an experiment to study the effect of additive noise on speech recognition systems / A. Varga, H. J. Steeneken // *Speech Communication*. – 1993. – Vol. 12, № 3. – P. 247–251. DOI: 10.1016/0167-6393(93)90095-3
18. Jaffard S. Wavelets: Tools for Science and Technology / S. Jaffard, Y. Meyer, R. D. Ryan. – Philadelphia: SIAM, 2001. – 256 p. DOI: 10.1063/1.1485591
19. Malvar H. S. Lapped transforms for efficient transform/subband coding / H. S. Malvar // *IEEE Transactions on Acoustics, Speech and Signal Processing*. – 1990. – Vol. 38, № 6. – P. 969–978. DOI: 10.1109/29.56057
20. Meyer F. G. Brushlets: A tool for directional image analysis and image compression / F. G. Meyer, R. R. Coifman // *Journal of Applied and Computational Harmonic Analysis*. – 1997. – Vol. 4, № 2. – P. 147–187. DOI: 10.1006/acha.1997.0208
21. Daubechies I. Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool / I. Daubechies, J. Lu, H.-T. Wu // *Journal of Applied and Computational Harmonic Analysis*. – 2011. – Vol. 30, № 2. – P. 243–261. DOI: 10.1016/j.acha.2010.08.002
22. Drugman T. Joint robust voicing detection and pitch estimation based on residual harmonics / T. Drugman, A. Alwan // *Proceedings of Interspeech*. – 2011. – P. 1973–1976.
23. Azarov E. Instantaneous pitch estimation based on RAPT framework / E. Azarov, M. Vashkevich, A. Petrovsky // *Proceedings of the European Signal Processing Conference*. – 2012. – P. 2787–2791.
24. Empirical Wavelet Transforms [Electronic resource]. – 2013. – Access mode: <http://www.mathworks.com/matlabcentral/fileexchange/42141-empirical-wavelet-transforms>

Статья поступила в редакцию 22.12.2014.

После доработки 12.01.2015.

Імамвердієв Я. Н.¹, Сухостат Л. В.²

¹Канд. техн. наук, зав. відділом, Інститут інформаційних технологій Національної Академії Наук Азербайджану, Баку, Азербайджан

²Науковий співробітник, Інститут інформаційних технологій Національної Академії Наук Азербайджану, Баку, Азербайджан

МЕТОД ОЦІНКИ ПЕРІОДУ ОСНОВНОГО ТОНУ З ЗАСТОСУВАННЯМ ЕМПІРИЧНОГО ВЕЙВЛЕТ ПЕРЕТВОРЕННЯ

Оцінка періоду основного тону мовного сигналу використовується в багатьох важливих областях застосування мовних технологій. Однак серед існуючих методів тільки деякі можуть працювати для випадку нелінійних і нестационарних сигналів. Основна причина полягає у тому, що методи виявлення періоду основного тону засновані на припущенні, що сам процес мовоутворення є лінійним, а мовні сигнали є локально стаціонарними. Вибір алгоритму для оцінки періоду основного тону завжди полягає у знаход-

женні компромісу між часом і роздільною здатністю за частотою, робастністю, затримкою й обчислювальною складністю. Метою даної роботи є розробка методу оцінки періоду основного тону на основі емпіричного вейвлет перетворення. Метод побудови сімейства адаптованих вейвлетів полягає у припущенні, що фільтри залежать від розташування інформації в спектрі аналізованого сигналу. Емпіричні вейвлети визначаються як смугові фільтри на кожному відрізьку мовного сигналу. Для визначення періоду основного тону як ознаки розглядається миттєва частота. Оператор поділу енергії Тігера-Кайзера використовується для її виділення. У роботі приводиться порівняння даного методу з іншими алгоритмами оцінки періоду основного тону.

Ключові слова: період основного тону, емпіричне вейвлет перетворення, оператор поділу енергії Тігера-Кайзера, внутрішня модова функція, миттєва частота.

Imamverdiyev Y. N.¹, Sukhostat L. V.²

¹PhD, Head of department, Institute of Information Technology of Azerbaijan National Academy of Sciences, Baku, Azerbaijan

²Research fellow, Institute of Information Technology of Azerbaijan National Academy of Sciences, Baku, Azerbaijan

PITCH PERIOD ESTIMATION METHOD USING EMPIRICAL WAVELET TRANSFORM

Pitch period evaluation of speech signal is used in many important applications of speech technology. However, among the existing methods only some can work in case of non-linear and non-stationary signals. The main reason is that the pitch detection methods are based on the assumption that speech production process is linear. Selection of pitch period estimation algorithm is always focuses on finding a compromise between time and frequency resolution, robustness, computational complexity and time delay. The aim of this paper is to develop a new method for estimating the pitch period based on empirical wavelet transformation. Method of constructing a family of adapted wavelets assumes that the filters depend on the information location in speech spectrum of the analyzed signal. Empirical wavelets are defined as bandpass filters for each segment of the speech signal. Instantaneous frequency characteristics are considered as pitch period detection features. Teager-Kaiser energy separation operator is used for its extraction. The comparison of this method with other pitch estimation algorithms is presented.

Keywords: pitch period, empirical wavelet transform, operator Teager-Kaiser energy operator, intrinsic mode function, instantaneous frequency.

REFERENCES

1. Rabiner L., Cheng M. J., Rosenberg A. E., McGonegal C. A. A comparative performance study of several pitch detection algorithms, *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1976, No. 5, pp. 399–417. DOI: 10.1109/TASSP.1976.1162846
2. Tan L. N., Alwan A. Multi-band summary correlogram-based pitch detection for noisy speech, *Speech Communication*, 2013, Vol. 55, No. 78, pp. 841–856. DOI: 10.1016/j.specom.2013.03.001
3. Ba H., Yang N., Demirkol I., Heinzelman W. BaNa: a hybrid approach for noise resilient pitch detection, *IEEE Statistical Signal Processing Workshop*, 2012, pp. 369–372. DOI: 10.1109/SSP.2012.6319706
4. De Cheveigne A., Kawahara H. Yin, a fundamental frequency estimator for speech and music, *Journal of the Acoustical Society of America*, 2002, Vol. 111, No. 4, pp. 1917–1930. DOI: 10.1121/1.1458024
5. Kasi K., Zahorian S. A. Yet another algorithm for pitch tracking, *Proceedings of the International Conference on Acoustics, Speech, and Signal*, Vol. 1, 2002, pp. 361–364. DOI: 10.1109/ICASSP.2002.5743729
6. Camacho A. SWIPE: a sawtooth waveform inspired pitch estimator for speech and music: thesis ... doctor of philosophy. Florida, University of Florida, 2007.
7. Gonzalez S., Brookes M. PEFAC – A Pitch Estimation Algorithm Robust to High Levels of Noise, *IEEE Transactions on Audio, Speech and Language Processing*, 2011, Vol. 22, No. 2, pp. 518–530. DOI: 10.1109/TASLP.2013.2295918
8. Boashash B. Estimating and interpreting the instantaneous frequency of a signal, *Proceedings of the IEEE*, 1992, Vol. 80, No. 4, pp. 520–538. DOI: 10.1109/5.135376
9. Maragos P., Kaiser J. F., Quatieri T. F. On amplitude and frequency demodulation using energy operators, *IEEE Transactions on Signal Processing*, 1993, Vol. 41, No. 4, pp. 1532–1550. DOI: 10.1109/78.212729
10. Abe T., Kobayashi T., Imai S. Harmonics tracking and pitch extraction based on instantaneous frequency, *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 1995, Vol. 1, pp. 756–759. DOI: 10.1109/ICASSP.1995.479804
11. Abe T., Honda M. /Sinusoidal model based on instantaneous frequency attractors, *IEEE Transactions on Audio, Speech and Language Processing*, 2006, Vol. 14, No. 4, pp. 1292–1300. DOI: 10.1109/TSA.2005.858545
12. Azarov E., Petrovsky A., Parfieniuk M. Estimation of the instantaneous harmonic parameters of speech, *Proceedings of the European Signal Processing Conference*, 2008.
13. Huang N. E., Shen Z., Long S. R., Wu M. L., Shih H. H., Zheng Q., Yen N. C., Tung C. C., Liu H. H. The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 1998, Vol. 454, No. 1971, P. 903–995. DOI: 10.1098/rspa.1998.0193
14. Gilles J. Empirical Wavelet Transform, *IEEE Transactions on Signal Processing*, 2013, Vol. 61, No. 16, pp. 3999–4010. DOI: 10.1109/TSP.2013.2265222
15. Vakman D. On the analytic signal, the Teager-Kaiser energy algorithm, and other methods for defining amplitude and frequency, *IEEE Transactions on Signal Processing*, 1996, Vol. 44, No. 4, pp. 791–797. DOI: 10.1109/78.492532
16. Chu W., Alwan A. Reducing f0 frame error of f0 tracking algorithms under noisy conditions with an unvoiced/voiced classification frontend, *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, 2009, pp. 3969–3972. DOI: 10.1109/icassp.2009.4960497
17. Varga A., Steeneken H. J. Assessment for automatic speech recognition: II. Noisex-92: a database and an experiment to study the effect of additive noise on speech recognition systems, *Speech Communication*, 1993, Vol. 12, No. 3, pp. 247–251. DOI: 10.1016/0167-6393(93)90095-3
18. Jaffard S., Meyer Y., Ryan R. D. Wavelets: Tools for Science and Technology, *Philadelphia: SIAM*, 2001, 256 p. DOI: 10.1063/1.1485591
19. Malvar H. S. Lapped transforms for efficient transform/subband coding, *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1990, Vol. 38, No. 6, pp. 969–978. DOI: 10.1109/29.56057
20. Meyer F. G., Coifman R. R. Brushlets: A tool for directional image analysis and image compression, *Journal of Applied and Computational Harmonic Analysis*, 1997, Vol. 4, No. 2, pp. 147–187. DOI: 10.1006/acha.1997.0208
21. Daubechies I., Lu J., Wu H.-T. Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool, *Journal of Applied and Computational Harmonic Analysis*, 2011, Vol. 30, No. 2, pp. 243–261. DOI: 10.1016/j.acha.2010.08.002
22. Drugman T., Alwan A. Joint robust voicing detection and pitch estimation based on residual harmonics, *Proceedings of Interspeech*, 2011, pp. 1973–1976.
23. Azarov E., Vashkevich M., Petrovsky A. Instantaneous pitch estimation based on RAPT framework, *Proceedings of the European Signal Processing Conference*, 2012, pp. 2787–2791.
24. Empirical Wavelet Transforms [Electronic resource], 2013, Access mode: <http://www.mathworks.com/matlabcentral/fileexchange/42141-empirical-wavelet-transforms>