

Метод адаптивного подавления шумов на основе эмпирического вейвлет преобразования

Людмила Сухостат

Институт Информационных Технологий НАНА

lsuhostat@hotmail.com

Аннотация— Обеспечение информационной безопасности в системах аутентификации пользователей является наиболее сложной и актуальной задачей в современном информационном обществе. Применение биометрических технологий являются одним из путей повышения точности работы таких систем. В практических приложениях распознавания диктора изменения акустической среды и каналов связи приводит к различным искажениям. Удаление шума имеет важное значение для снижения помех в речевом сигнале, для повышения точности и надежности. В статье предлагается метод на основе эмпирического вейвлет преобразования, который был проверен на речевых сигналах с различным уровнем шума.

Ключевые слова— речевые признаки; вейвлеты; эмпирическое вейвлет преобразование; внутренняя модовая функция

I. ВВЕДЕНИЕ

Искажение канала является серьезной проблемой для систем верификации и идентификации говорящего, так как даже незначительное количество искажений в речи может вызвать изменение уникальных особенностей в голосе человека и уничтожение необходимой информации.

В настоящее время разработано большое количество методов распознавания диктора на основе нескольких подходов, которые показывают обнадеживающие результаты по идентификации и верификации личности по голосу. Однако вопрос о создании систем идентификации диктора, обладающих высокой точностью остается открытым.

Основной целью исследований в области распознавания говорящего является создание методов и алгоритмов, повышающих точность распознавания, сохраняя при этом приемлемые показатели по вычислительной трудоемкости.

Системы, основанные на речевых технологиях, имеют существенные преимущества по сравнению с другими биометрическими системами, поэтому к ним предъявляются более высокие требования обеспечения безопасности и требуется разработка новых методов и алгоритмов для повышения их точности.

В настоящее время широкое распространение получили вейвлет преобразования, которые применяются к обработке различных сигналов для выделения характерных особенностей.

Вейвлеты адаптируются к сигналу, характеризуя его локальные свойства. В отличие от других известных методов работы с сигналами, таких как оконное

преобразование Фурье, вейвлеты обладают рядом преимуществ:

- наличие окна переменного размера для проведения гибкого анализа;
- предоставление частотно-временной информации о сигнале;
- применение банка полосовых фильтров.

Известный подход, называемый «синхронное сжатие» (synchrosqueezing), предложенный в работах [1, 2], основан на выборе соответствующих вейвлетов. Они удаляют незначительные вейвлет коэффициенты (во времени и масштабе) учитывая пороговое значение соответствующей энергии сигнала в этой части.

Другие недавние работы для достижения той же цели включают эмпирическое вейвлет преобразование (Empirical Wavelet Transform, EWT) для построения адаптивного вейвлет-базиса для разложения заданного сигнала на адаптивные «полосы» [3]. Эта модель основана на робастной предварительной обработке для обнаружения пиков, а затем выполняет сегментацию спектра, основываясь на обнаруженных максимумах, и формирует соответствующий банк вейвлет фильтров. Банк фильтров обладает гибкостью на спектральных перекрытиях.

Данный адаптивный метод приводит к разложению сигнала на его основные моды. Метод сочетает в себе силу вейвлет формализма и адаптируемость эмпирической модовой декомпозиции (Empirical Mode Decomposition, EMD). Действительно, современные модели разложения в основном ограничены:

- их алгоритмической природе не хватает математической теории (EMD);
- рекурсивный отсев в большинстве методов, который не позволяет обратную коррекцию ошибок;
- неспособность должным образом справиться с шумом;
- жесткие граничные пределы вейвлет подходов.

EWT является одним из адаптивных методов используемых, где базис находится в зависимости от информационного содержания в сигнале. Метод имеет большое преимущество в выделении стационарных и нестационарных компонентов из сигнала. Здесь мода представлена в качестве компонентов AM-FM [4]. Производится сегментация спектра Фурье и небольшая фильтрация применяется для обнаружения опорных

границ. Ортонормированный базис формируется на основе информации, представленной в сигнале.

В данной работе, учитывая вышеизложенное, предлагается новый подход извлечения признаков речевых сигналов на основе EWT.

II. НОВЫЙ ПОДХОД ИЗВЛЕЧЕНИЯ ПРИЗНАКОВ РЕЧЕВЫХ СИГНАЛОВ НА ОСНОВЕ ЭМПИРИЧЕСКОГО ВЕЙВЛЕТ ПРЕОБРАЗОВАНИЯ

Исходя из этого применяется EWT и обратное преобразование для анализа сигнала. EWT позволяет отделять как аппроксимирующие, так и детализирующие коэффициенты [5]. Он основан на априорной фильтрации и имеет сильный математический базис. В силу шумоподавления EWT широко применяется к нестационарным сигналам.

В первую очередь определяется спектр Фурье речевого сигнала на отрезке $[0, \pi]$, а затем он делится на N различных сегментов. Границы между сегментами обозначаются через $\omega_n, n = \overline{0, N}$ (где $\omega_0 = 0$ и $\omega_N = \pi$). Находится максимум или наивысший пик сигнала. В середине между двумя максимумами рассматривается средняя точка. Строится банк фильтров. Выходной фильтр свертывается с исходным речевым сигналом с целью получения внутренних модовых функций (Intrinsic Mode Functions, IMF). Вначале получают зашумленные IMF, которые подвергаются фильтрации, чтобы очистить речевой сигнал от шума.

Сегментация осуществляется для получения «жестких» фреймов, чтобы избежать потери информации. Для этого должно удовлетворяться следующее условие:

$$\gamma < \min_n \left(\frac{\omega_{n+1} - \omega_n}{\omega_{n+1} + \omega_n} \right).$$

В работе эмпирическая масштабируемая функция и эмпирические вейвлеты рассматриваются и определяются следующим образом

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

и

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

где $0 < \gamma < 1$ для всех $n > 0$.

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

$$\beta(x) = \begin{cases} 3x^2 - 2x^3, & \text{если } 0 \leq x \leq 1 \\ 0, & x < 0 \\ 1, & 1 < x \end{cases}, \quad (3)$$

если $\beta(x) = 1 - \beta(1-x), \forall x \in [0,1]$.

Теперь можно определить эмпирическое вейвлет преобразование $W_f^\varepsilon(n, t)$ также как и в случае классического вейвлет преобразования

$$W_f^\varepsilon(n, t) = \langle f, \psi_n \rangle = \int f(\tau) \overline{\psi_n(\tau - t)} d\tau = \hat{f}(\omega) * \overline{\hat{\psi}_n(\omega)}, \quad (4)$$

а аппроксимирующие коэффициенты представить в виде скалярных произведений с масштабируемой функцией

$$W_f^\varepsilon(0, t) = \langle f, \phi_1 \rangle = \int f(\tau) \overline{\phi_1(\tau - t)} d\tau = \hat{f}(\omega) * \overline{\hat{\phi}_1(\omega)}, \quad (5)$$

где $\hat{\psi}_n(\omega)$ и $\hat{\phi}_1(\omega)$ определяются из уравнений (1) и (2) соответственно. Обратное преобразование принимает вид

$$\begin{aligned} f(t) &= W_f^\varepsilon(0, t) * \phi_1(t) + \sum_{n=1}^N W_f^\varepsilon(n, t) * \psi_n(t) = \\ &= \hat{W}_f^\varepsilon(0, \omega) * \hat{\phi}_1(\omega) + \sum_{n=1}^N \hat{W}_f^\varepsilon(n, \omega) * \hat{\psi}_n(\omega). \end{aligned} \quad (6)$$

Функция IMF f_k определяется следующим образом:

$$f_0(t) = W_f^\varepsilon(0, t) * \phi_1(t), \quad (7)$$

$$f_k(t) = W_f^\varepsilon(k, t) * \psi_k(t). \quad (8)$$

После получения IMF применяется дискретный алгоритм разделения энергии (Discrete Energy Separation Algorithm, DESA) [6], который обладает низкой вычислительной сложностью и быстро работает на реальных сигналах.

Пусть $d^m(n)$ - значение IMF для каждого фрейма при $n = 1, \dots, N$ и $m = 1, \dots, M_x$, где M_x обозначает число мод, на которые $x(t)$ разбивается.

Затем можно применить дискретный оператор Тигера

$$\Psi[d^m(n)] = (d^m(n))^2 - d^m(n-1)d^m(n+1), n = \overline{2, N-1}. \quad (9)$$

если $d^m(n)$ дискретный косинусный с постоянной амплитудой A и частотой ω , $d^m(n) = A \cos(\Omega n + \theta)$ при $\Omega = \omega T$ и T - период дискретизации, то

$$\Psi[d^m(n)] = A^2 \omega^2 \left(\frac{\sin \Omega}{\Omega} \right)^2. \quad (10)$$

В качестве параметров были рассмотрены мгновенная частота $\Omega(n)$ и мгновенная амплитуда $a(n)$:

$$\Omega(n) = \arccos\left(1 - \frac{\Psi[y(n)] + \Psi[y(n+1)]}{4\Psi[d^m(n)]}\right), \quad (11)$$

$$|a(n)| = \frac{\Psi[d^m(n)]}{\sqrt{1 - \left(1 - \frac{\Psi[y(n)] + \Psi[y(n+1)]}{4\Psi[d^m(n)]}\right)^2}}, \quad (12)$$

где $y(n) = d^m(n) - d^m(n-1)$ для $n = 2, \dots, N$.

Мгновенная частота и мгновенная амплитуда объединяются для получения средневзвешенной по амплитуде кратковременной оценки мгновенной частоты F_{mean} для каждого речевого сигнала [7] согласно

$$F_{mean} = \frac{\sum_{n=2}^N \Omega(n)a(n)}{\sum_{n=2}^N a(n)} \quad (13)$$

Выражение (11) обеспечивает более точную оценку и более робастно к низкой энергии и полосам зашумленных частот.

III. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

Для проведения экспериментальных исследований была применена система компьютерного моделирования MATLAB, в состав которой входит дополнение Wavelet Toolbox для анализа речевых сигналов. Сигнал был получен из базы данных Keele [8]. С помощью разработанного метода на основе EWT при различных процентах уровня шума вычисляется SNR сигнала. Внешний вид масштабируемой функции показан на рис. 1.

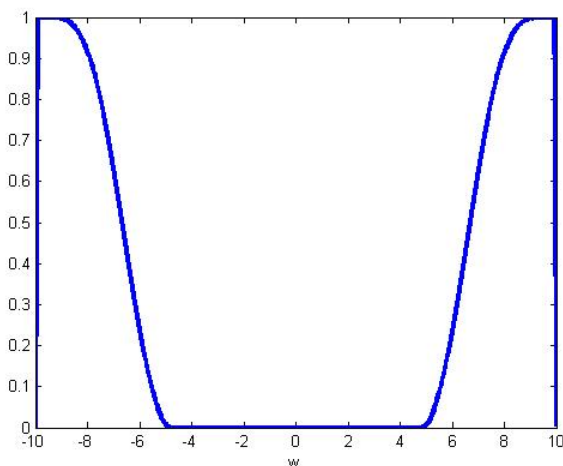


Рис 1. Вид масштабируемой функции

Видно, что с уменьшением ширины фильтра, т.е. с уменьшением масштабирующего параметра, одновременно увеличивается амплитуда. Т.о. полученная масштабируемая функция обладает высокой чувствительностью к кратковременным высокочастотным изменениям сигнала, что способствует снижению уровня шума.

ЗАКЛЮЧЕНИЕ

При обработке речевых сигналов удаление шума является одной из важных проблем.

Единственный способ их преодоления – это иметь модели, которые постоянно адаптируются к изменениям речевого сигнала.

Учитывая вышеизложенное, предложенный метод на основе EWT показал достаточно обнадеживающие результаты при различных уровнях шума, которые позволяют выделить ряд направлений для дальнейших исследований: подбор базовых функций и типов преобразований для повышения точности работы предложенного метода, а также его применение в области распознавания личности по голосу.

БИБЛИОГРАФИЯ

- [1] I. Daubechies, J. Lu, and H.-T. Wu, "Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 243–261, 2010.
- [2] H.-T. Wu, P. Flandrin, and I. Daubechies, "One or Two Frequencies? the Synchrosqueezing Answers," *Advances in Adaptive Data Analysis*, vol. 3, no. 1-2, pp. 29–39, 2011.
- [3] J. Gilles, "Empirical Wavelet Transform," *IEEE Transactions on Signal Processing*, vol. 61, no. 16, pp. 3999–4010, 2013.
- [4] R.S. Holambe, M.S. Deshpande, "Noise robust speaker identification: using nonlinear modeling", *Forensic Speaker Recognition*, pp. 153–182, 2012.
- [5] Я.Н. Имамвердиев, Л.В. Сухостат, "Разработка робастного метода извлечения речевых признаков на основе эмпирического вейвлет преобразования", *Информационные технологии*, no. 1, стр. 19-23, 2015.
- [6] G. Schlotthauer, M.E. Torres. and H.L. Rufiner, "A new algorithm for instantaneous F0 speech extraction based on Ensemble Empirical Mode Decomposition", *Proc. of 17th European Signal Processing Conf.*, pp. 2347–2351, 2009.
- [7] S. Chhabra, R. Bajaj, R.B. Pachori, and R.N. Biswas, "Features based on Fourier-Bessel expansion for application of speaker identification system", *Proc. Of Indian Conf. for Academic Research by Undergraduate Students*, 26-28 March, 2010, IIT Kanpur, India.
- [8] F. Plante, G. Meyer, W.A. Ainsworth, "A pitch extraction reference database", *Proc. of Eurospeech*, pp. 837-840, 1995.