

# Детектирование активности голоса основанное на автокорреляционной функции с использованием вейвлет преобразование и оператора энергии Тигера

---

Voice Activity Detection(VAD) Based on Auto-Correlation  
Function Using Wavelet Transform and Teager Energy  
Operator

# Что такое и Зачем нужно детектирование

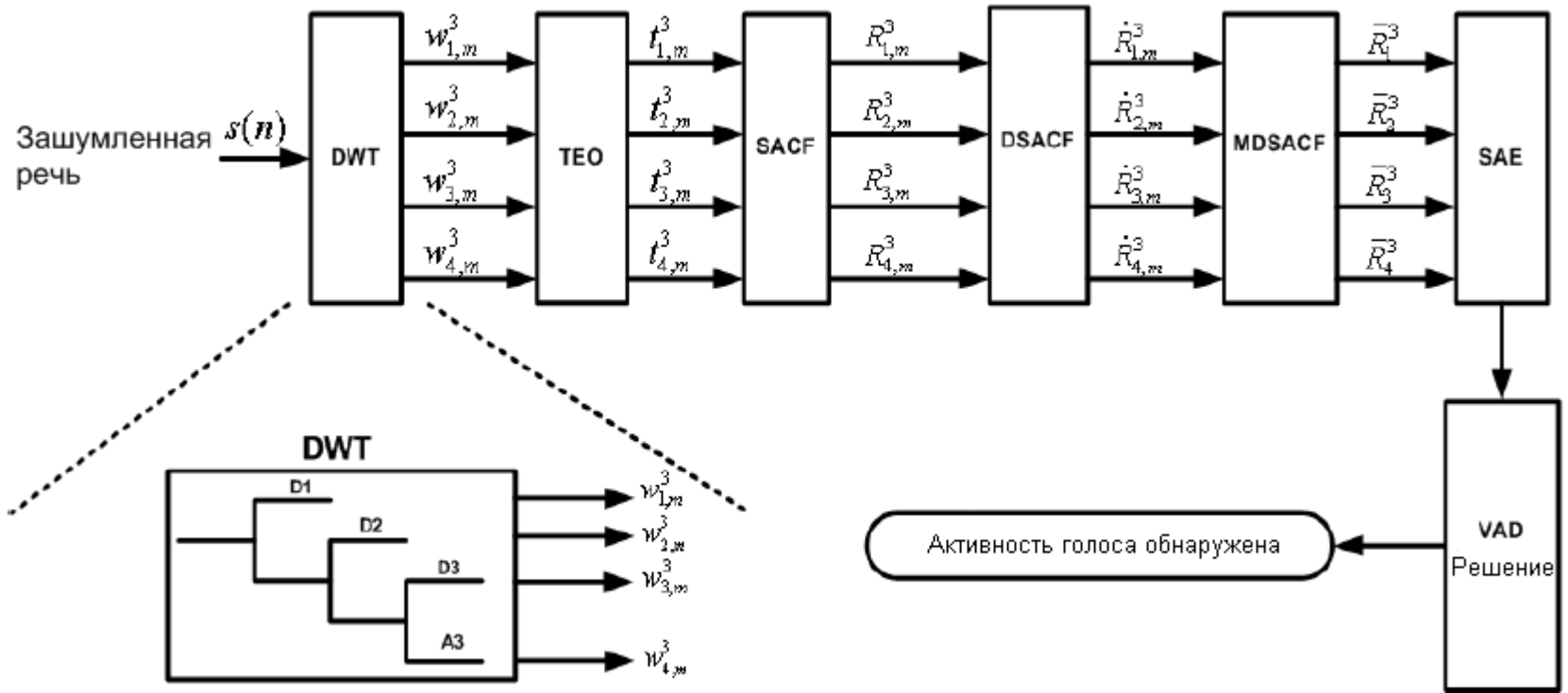
- VAD предоставляет возможность различить речь и шум, является основной частью различных систем речевой коммуникации, таких как кодирование речи, распознавание, hand-free телефонии и устранения эха.

- Пример:

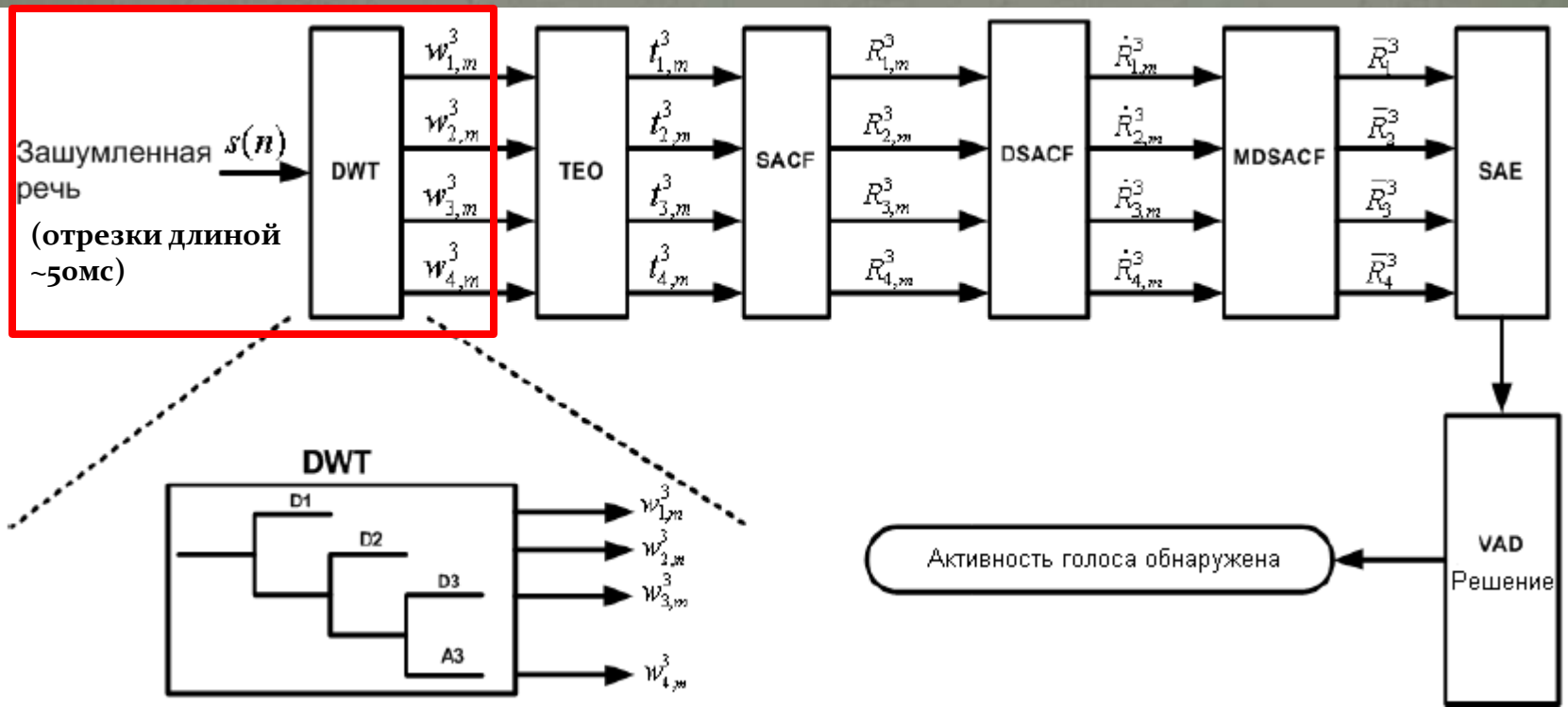
VAD схема используется в коммуникациях основанных на GSM для удлинения времени работы батареи клиента путём прерывания передачи сигнала при обнаружении пауз в речи.

Для сжатия речевых сигналов и уменьшения трафика в VoIP системах.

# Данный алгоритм

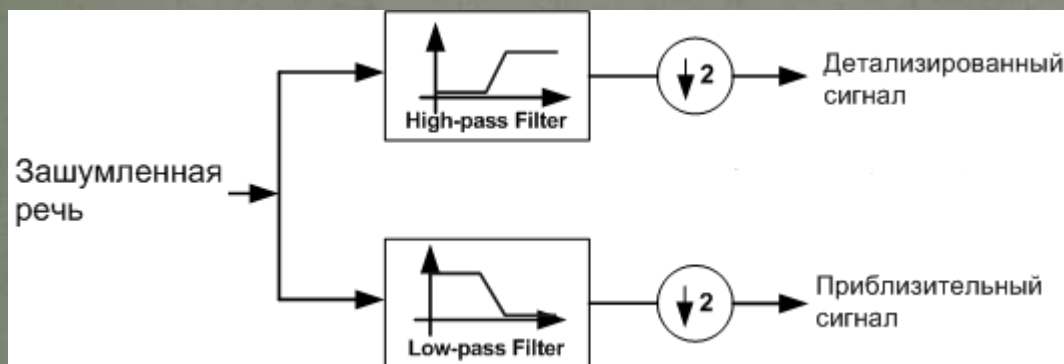



# Данный алгоритм



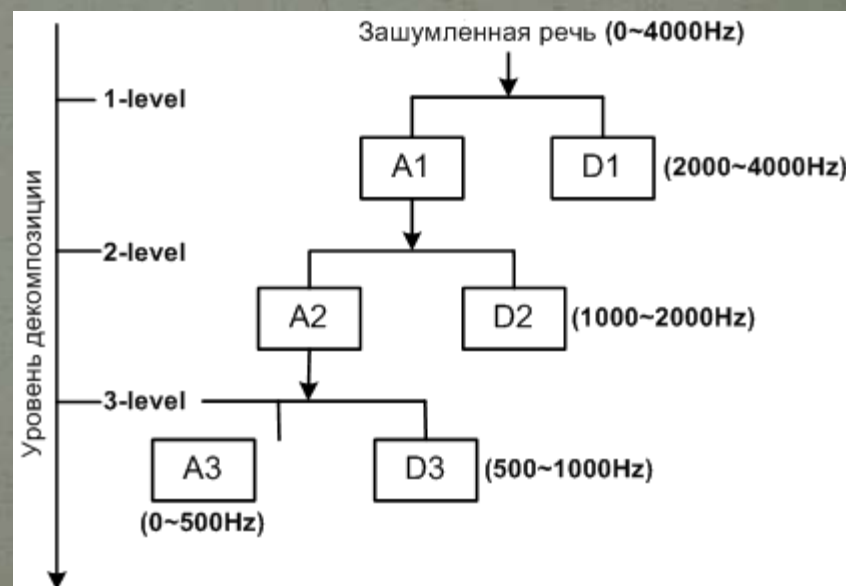


# Дискретное вейвлет преобразование (*Discrete wavelet transform* – DWT)

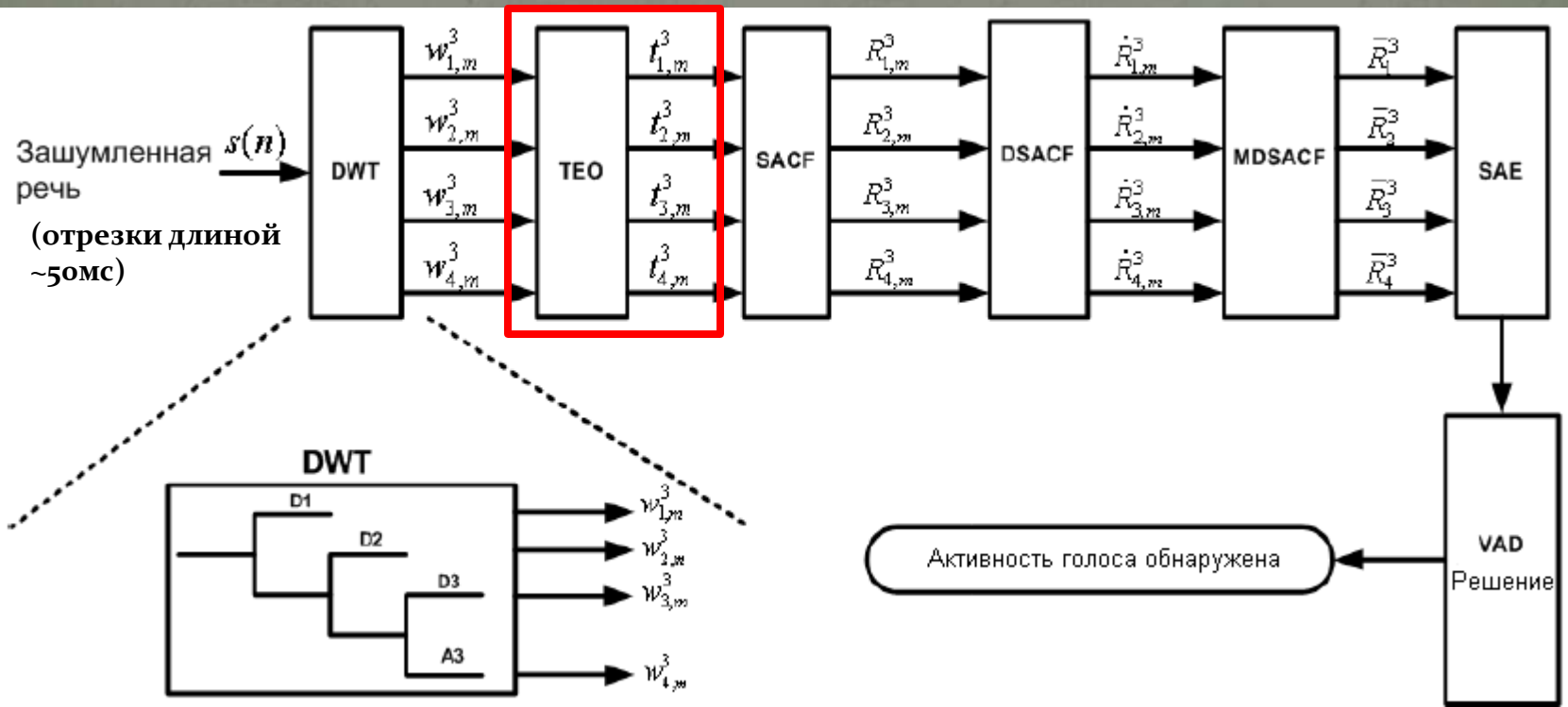


 - оператор уменьшения выборки в 2 раза

Трехуровневая  
вейвлет декомпозиция  
с использованием фильтров



# Данный алгоритм



# Опертор энергии Тигера

(Teager Energy Operator – TEO)

$$\psi_c[s(t)] = [\dot{s}(t)]^2 - s(t)\ddot{s}(t)$$

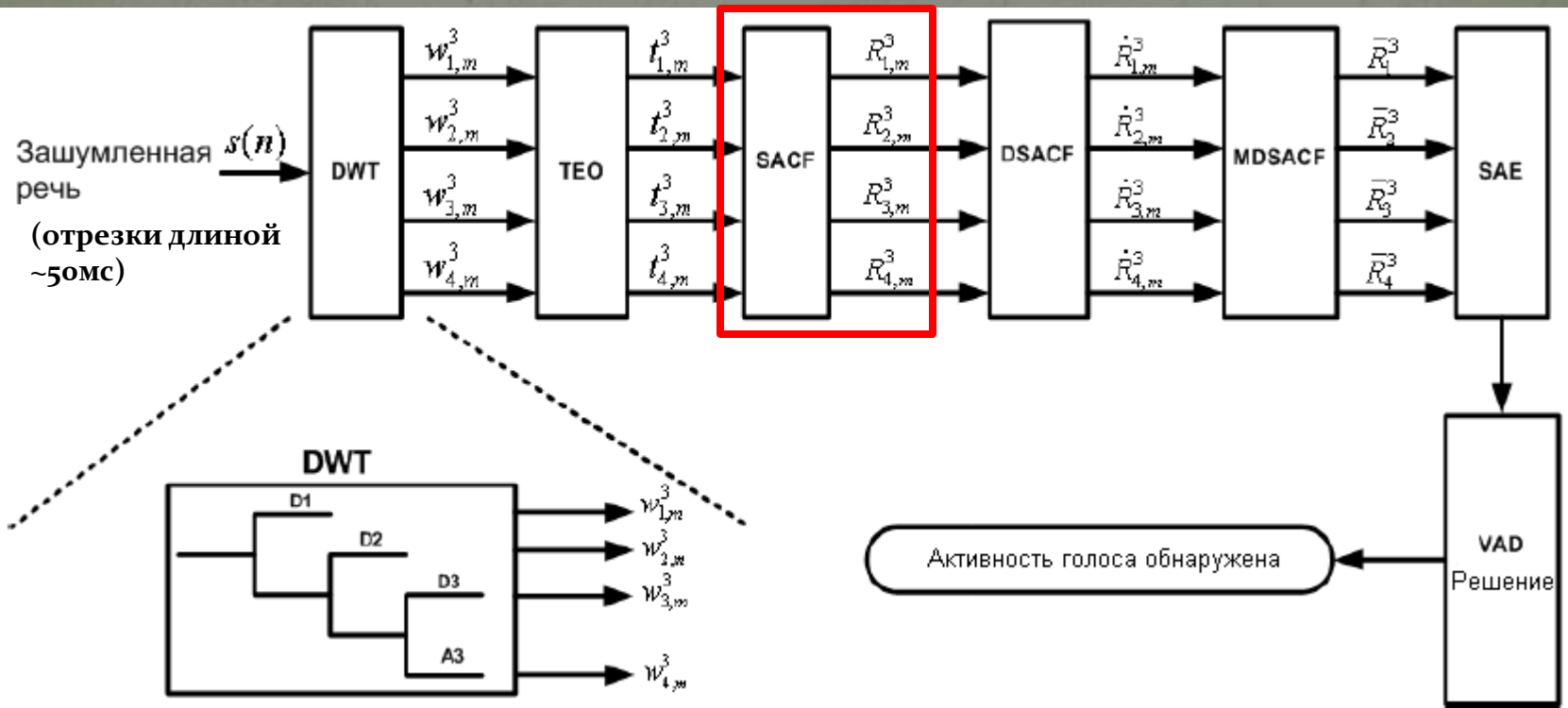
$$\dot{s} = ds/dt$$



$$\psi_d[s(n)] = s(n)^2 - s(n+1)s(n-1)$$

Эксперименты показывают что значение энергии Тигера для речи много больше чем для шума.

# Данный алгоритм





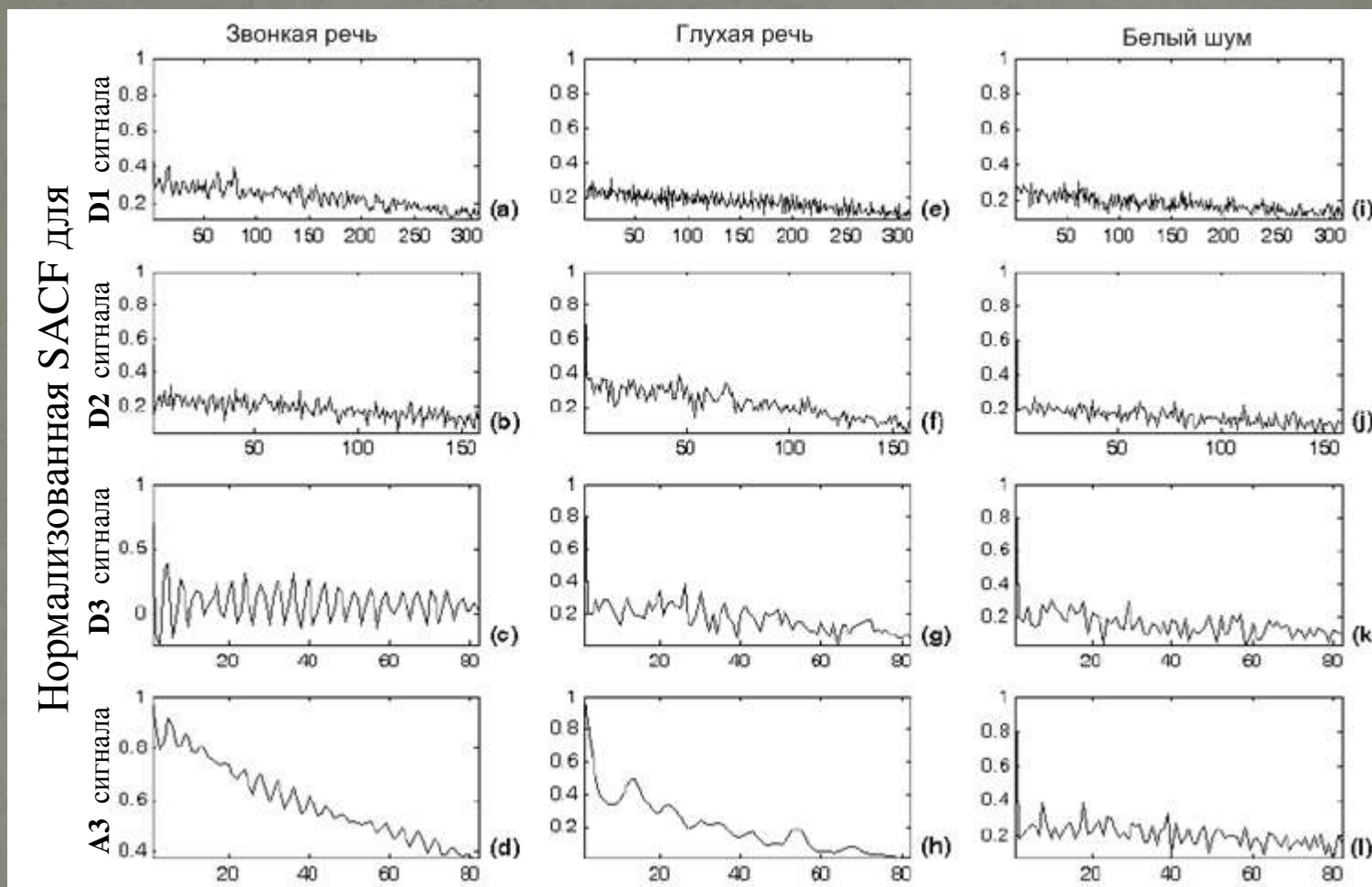
## Полосовая автокорреляционная функция (Subband Auto-Correlation Function -SACF)

$$R(k) = \sum_{n=0}^{p-k} s(n)s(n+k), \quad k = 0, 1, \dots, p$$

Автокорреляционная функция

Полосовая – применяется к частотным полосам сигнала, полученным при помощи вейвлет преобразования.

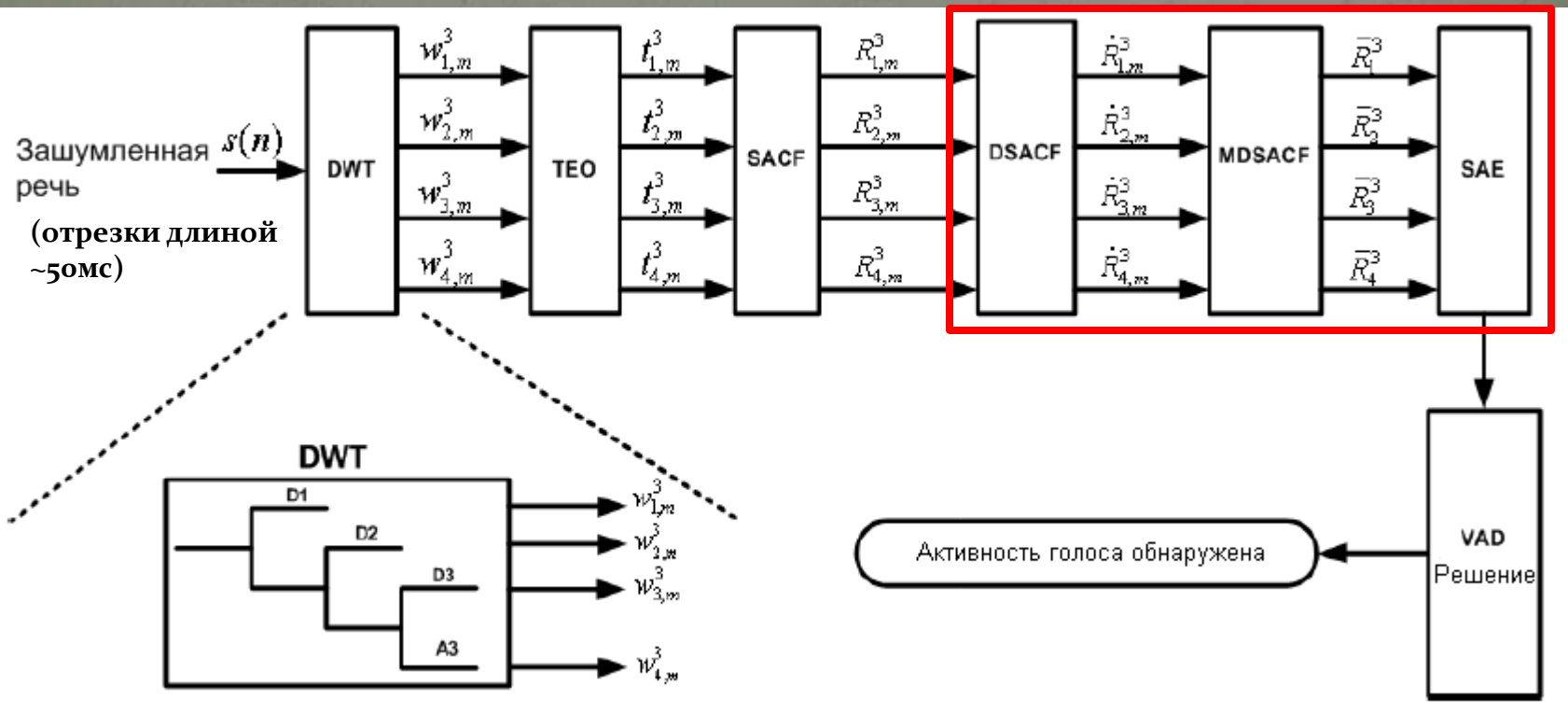
# Полосовая автокорреляционная функция (Subband Auto-Correlation Function -SACF)



Нормализованная -  $R(0) = 1$

D1, D2, D3, A3 – полосы сигнала после DWT

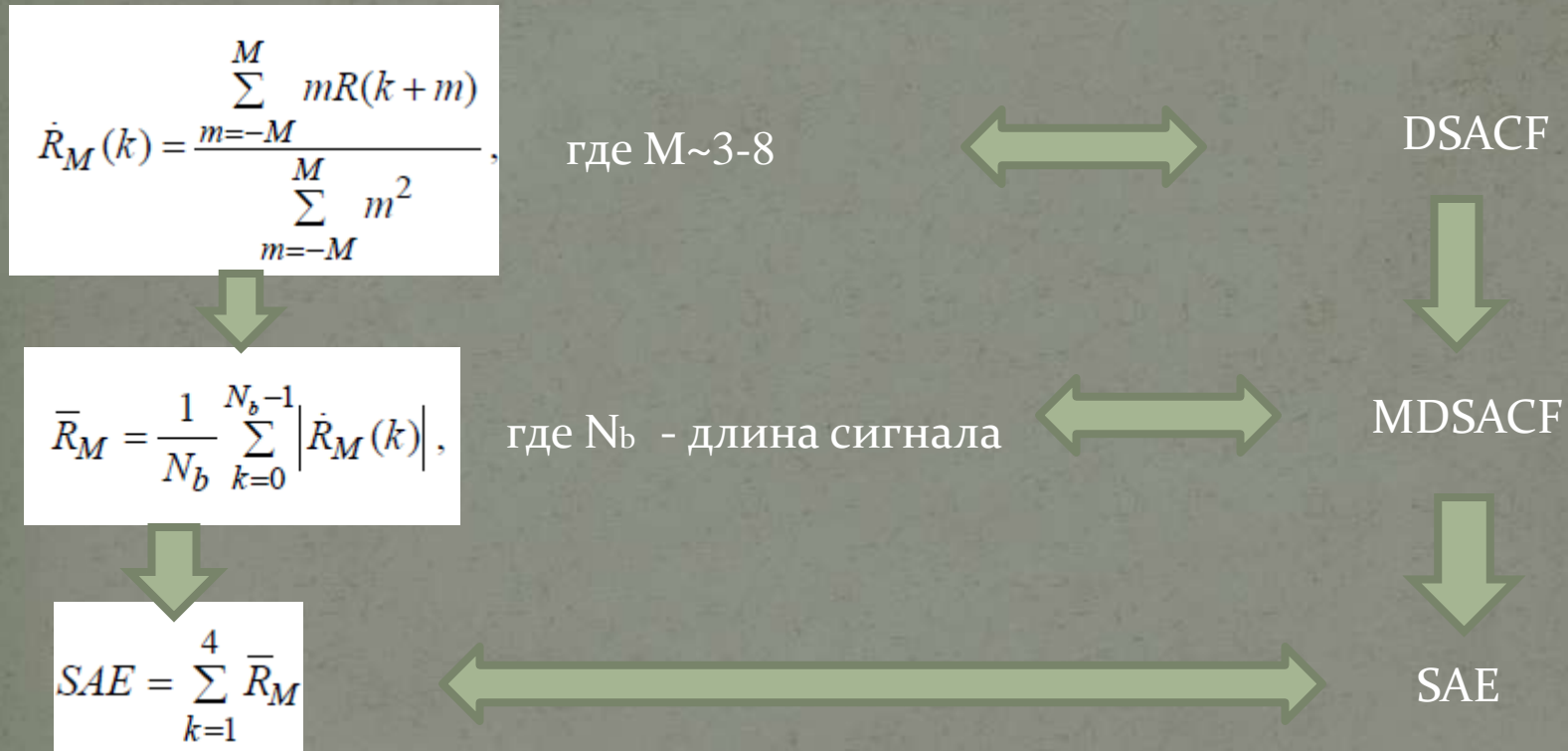
# Данный алгоритм



Средняя дельта полосовая автокорреляционная функция  
(Mean Delta Subband Signal Auto-Correlation Function -MDSACF)

и

Огибающая речевой активности (Speech Activity Envelope -SAE)

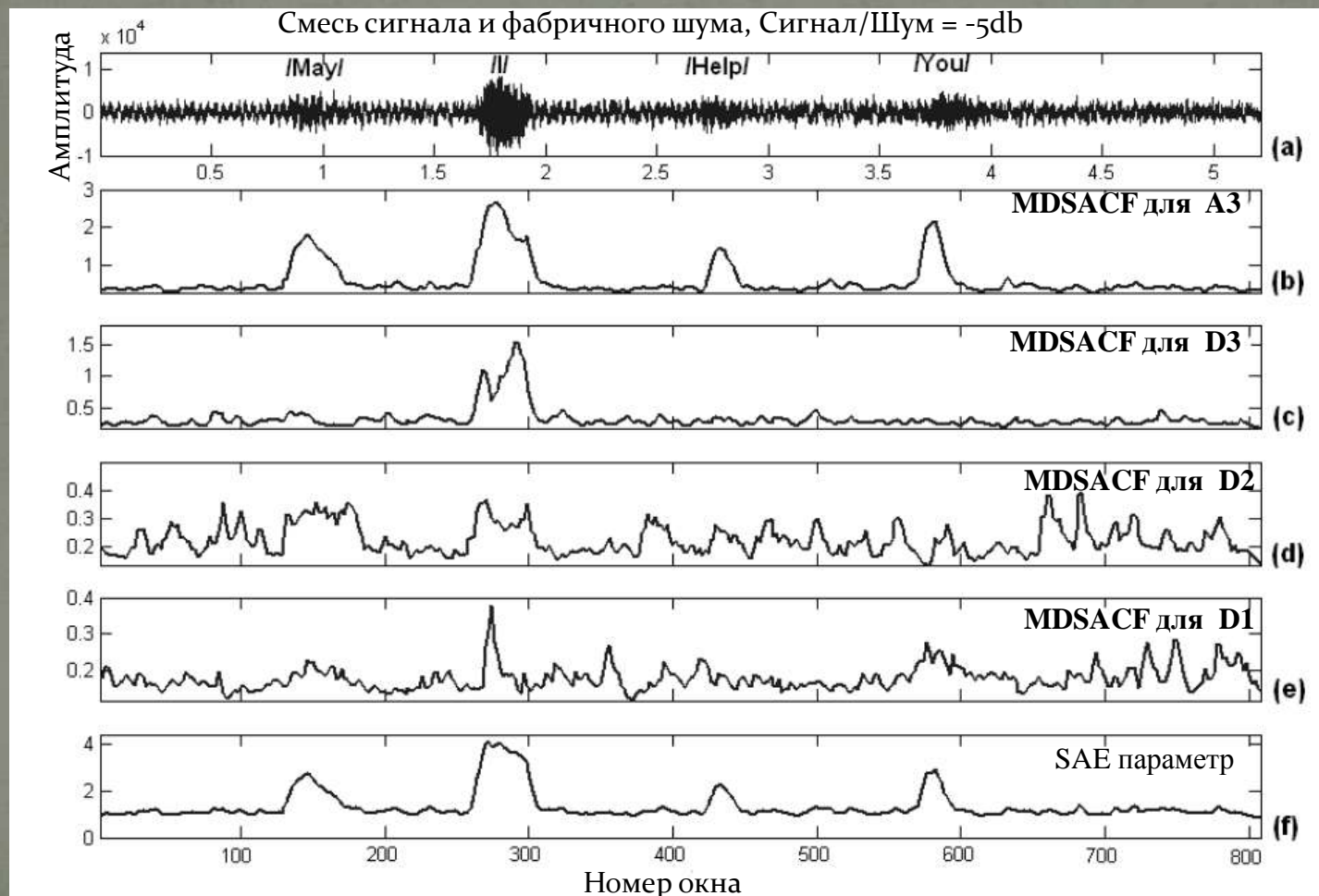




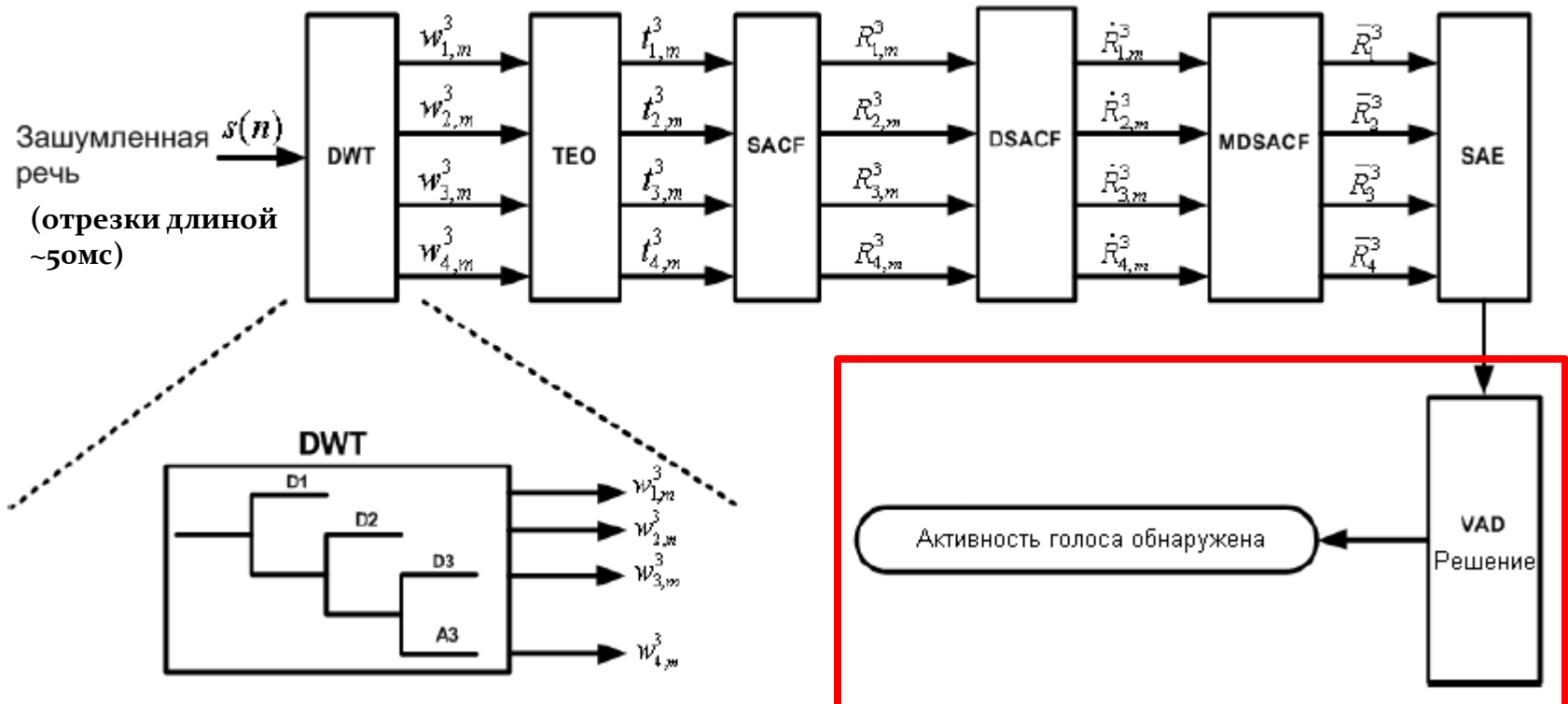
# Средняя дельта полосовая автокорреляционная функция (Mean Delta Subband Signal Auto-Correlation Function -MDSACF)

И

## Огибающая речевой активности(Speech Activity Envelope -SAE)



# Данный алгоритм



# Принятие решения Речь-Шум

$$T_s = \mu_n + \alpha_s \cdot \sigma_n, \quad - \text{адаптивный порог речи}$$

$$T_n = \mu_n + \beta_n \cdot \sigma_n, \quad - \text{адаптивный порог шума}$$

$$\mu_n(t) = \gamma \cdot \mu_n(t-1) + (1-\gamma) \cdot SAE(t),$$

$$\sigma_n(t) = \sqrt{[SAE_{buffer}^2]_{mean} - [\mu_n(t)]^2},$$

$$\alpha_s \sim 5$$

$$\beta_n \sim -1$$

$$\gamma \sim 0.95$$

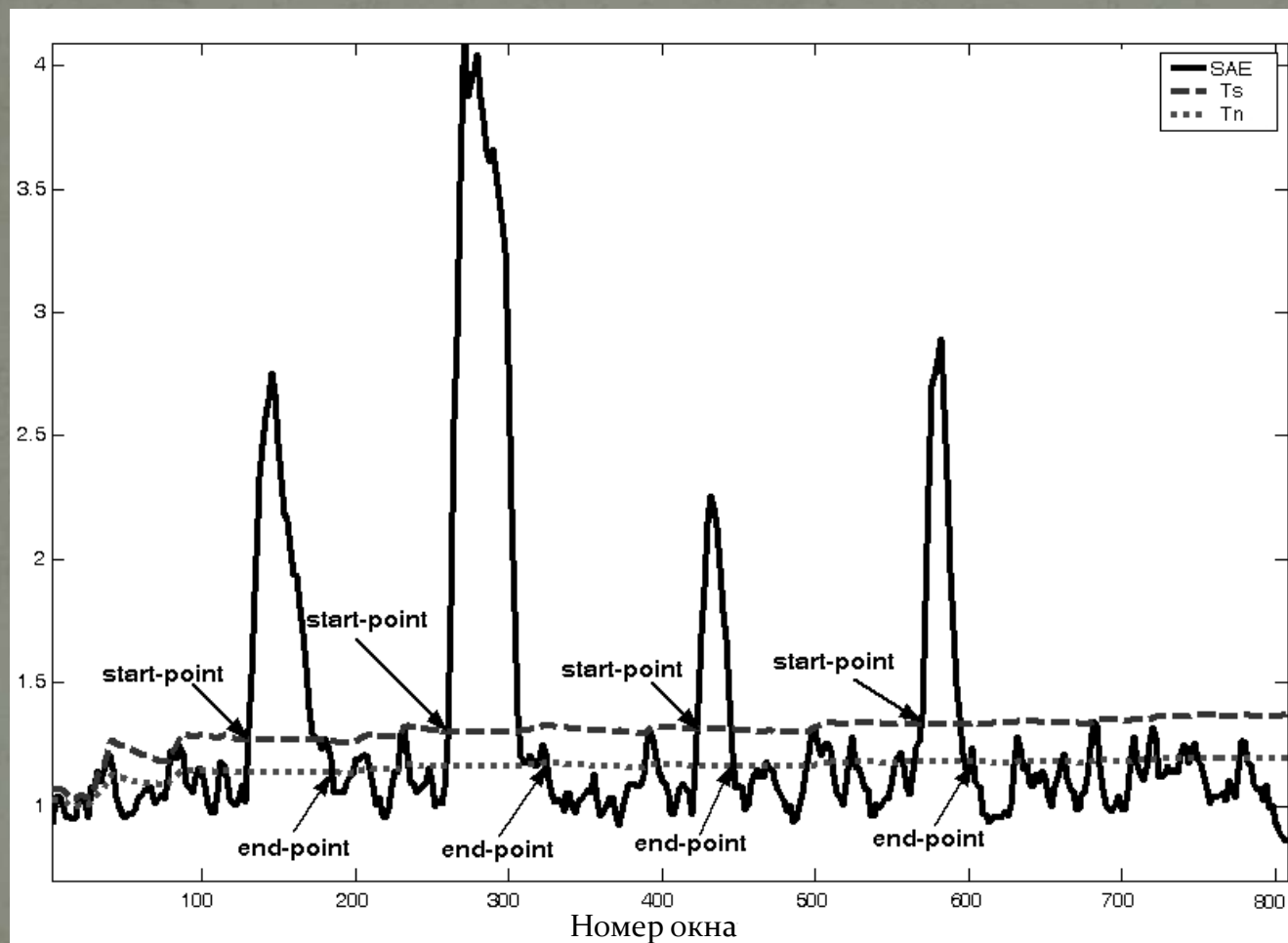
$$[SAE_{buffer}^2]_{mean}(t) = \gamma \cdot [SAE_{buffer}^2]_{mean}(t-1) + (1-\gamma) \cdot SAE(t)^2.$$

if  $(SAE(t) > T_s)$   $VAD(t)=1$

else if  $(SAE(t) < T_n)$   $VAD(t)=0$ ; - принятие решения

else  $VAD(t)=VAD(t-1)$ .

# Принятие решения Речь-Шум





# Результаты. Качество

| Шумовые условия |           | Вероятность верного решения(%) |              |              | Вероятность неверного решения(%) |            |              |
|-----------------|-----------|--------------------------------|--------------|--------------|----------------------------------|------------|--------------|
| Тип             | С/Ш (dB)  | Данный VAD                     | Chen's VAD   | G.729B VAD   | Данный VAD                       | Chen's VAD | G.729B VAD   |
| Автомобильный   | <b>30</b> | 99.1                           | 97.3         | 92.1         | 6.2                              | 6.9        | 7.3          |
|                 | <b>10</b> | 97.3                           | 96.1         | 86.5         | 8.6                              | 9.3        | 16.3         |
|                 | <b>-5</b> | 92.6                           | 93.5         | 72.3         | 10.5                             | 10.9       | 21.5         |
| Заводской       | <b>30</b> | 96.9                           | 97.2         | 96.9         | 7.6                              | 10.3       | 9.1          |
|                 | <b>10</b> | 93.1                           | 94.1         | 82.3         | 8.8                              | 13.2       | 18.9         |
|                 | <b>-5</b> | 87.2                           | 85.6         | 70.7         | 10.9                             | 15.4       | 26.4         |
| Белый шум       | <b>30</b> | 99.1                           | 97.2         | 98.4         | 1.3                              | 1.9        | 2.0          |
|                 | <b>10</b> | 98.5                           | 98.1         | 86.3         | 1.5                              | 1.8        | 3.6          |
|                 | <b>-5</b> | 93.2                           | 92.9         | 60.5         | 1.6                              | 2.3        | 3.3          |
| Среднее         |           | <b>95.22</b>                   | <b>94.67</b> | <b>82.89</b> | <b>6.33</b>                      | <b>8</b>   | <b>12.04</b> |

# Результаты. Скорость вычисления

| Тип VAD    | Время вычисления свойства | Время принятия решения |
|------------|---------------------------|------------------------|
| G.729B     | 0.048 s                   | 0.023 s                |
| Chen'VAD   | 4.126 s                   | 0.098 s                |
| Данный VAD | 0.23 s                    | 0.12 s                 |

Всё.

Ссылка:

**Bing-Fei Wu, Kun-Ching Wang** “Voice Activity Detection Based on Auto-Correlation Function Using Wavelet Transform and Teager Energy Operator”  
*Computational Linguistics and Chinese Language Processing Vol. 11, No. 1, March 2006, pp. 87-100*