

С.Д. ТИУНОВ, Р.В. МЕЩЕРЯКОВ, Д.В. ЧЕРНЫХ
**ОПТИМИЗАЦИЯ ВЫЧИСЛЕНИЯ ОДНОВРЕМЕННОЙ
МАСКИРОВКИ РЕЧЕВОГО СИГНАЛА**

Тиунов С.Д., Мещеряков Р.В., Черных Д.В. **Оптимизация вычисления одновременной маскировки речевого сигнала.**

Аннотация. В настоящей статье предлагается оптимизация вычисления одновременной маскировки речевого сигнала для реализации в задачах реального времени. Показана трудоемкость циклической свертки для итерационного алгоритма Тоома-Кука длины 4 и алгоритма на основе БПФ. Делается вывод об эффективности предлагаемых решений по вычислительной сложности и по объемам занимаемой памяти.

Ключевые слова: речевой сигнал, одновременная маскировка, циклическая свертка, оптимизация.

Tiunov S.D., Meshcheryakov R.V., Chernykh D.V. **Optimization of calculation of frequency masking of speech signal.**

Abstract. Paper proposes optimization of calculation of frequency masking of speech signal for use in real time applications. The complexity of circular convolution is shown for iterative Toom-Cook algorithm with length 4 and an algorithm based on the FFT. The conclusion about the efficiency of the proposed solutions is drawn on computation complexity and memory.

Keywords: speech signal, frequency masking, circular convolution, optimization.

1. Введение. Обработка речи включает в себя различные затратные операции над звуковым сигналом [1, 2]. Вместе с тем, использование новых подходов и методов обработки сигнала требует значительных вычислительных ресурсов и не может быть проведено в реальном масштабе времени [2-5]. Это связано с одной стороны с тем, что при разработке новых алгоритмов для обработки речи на первом месте стоит вопрос точности, а не вопрос вычислительной эффективности. Вместе с тем некоторые приложения речевых технологий подразумевают обработку сигнала в реальном времени. Данная статья предлагает эффективный численный метод для вычисления одновременной маскировки из модели слуховой системы [2, 6], что позволяет применять ее в программном комплексе для анализа речевого сигнала в норме и при патологии [7-10].

Одновременная маскировка является особенностью восприятия звука слуховой системой человека. Эта особенность заключается в том, что из двух близких частотных составляющих ухом будет воспринята только та, которая имеет большую интенсивность [10]. Моделирование эффекта одновременной маскировки позволяет выделить в спектрограмме сигнала наиболее важные (информативные) составляющие с точки зрения слуховой системы человека. Моделирование одновременной маскировки и связанные алгоритмы подробно описаны в [11].

Алгоритм одновременной маскировки может быть представлен в виде функции m , которая принимает на вход матрицу спектрограммы S_{KN} и дает на выходе двоичную матрицу одновременной маскировки M_{KN} (единица означает, что данная составляющая воспринимается слуховой системой). При этом для своей работы алгоритм использует матрицу H_{KL} – маскирующую матрицу (см. рисунок 1).



Рис. 1. Вычисление одновременной маскировки речевого сигнала как черный ящик

Формула для вычисления функционала m может быть записана следующим образом:

$$\forall k = 0 \dots K - 1, \forall n = 0 \dots N - 1:$$

$$M_{kn} = \left(\sum_{l=0}^{L-1} |S_{k'n}|^2 \cdot H_{kl} \right) \cdot b > |S_{kn}|^2, \quad (1)$$

$$\text{где } k' = \begin{cases} 0, & \text{если } k + l - L/2 < 0 \\ k + l - L/2, & \text{если } 0 \leq k + l - L/2 \leq K - 1 \\ K - 1, & \text{если } K - 1 < k + l - L/2. \end{cases}$$

$$b = \text{const}$$

Предполагая, что квадрат модуля спектра $|S_{kn}|^2$ вычислен заранее, запишем вычислительную сложность (количество умножений M и количество сложений A) «наивного» алгоритма одновременной маскировки (т. е. алгоритма, использующего формулы (1) непосредственно):

$$M(K, N, L) = N \cdot K \cdot (L - 1), \quad (2)$$

$$A(K, N, L) = N \cdot K \cdot (L + 1), \quad (3)$$

$$M + A = 2 \cdot N \cdot K \cdot L. \quad (4)$$

Рассчитаем требуемую вычислительную мощность для наивного алгоритма при различных значениях параметров (таблица 1). Для получения значения в FLOPS (количество вещественных операций в

секунду) необходимо принять $N = f_q$, где f_q – частота дискретизации сигнала.

Таблица 1. Расчеты мощности вычислительной машины для расчета одновременной маскировки по наивному алгоритму

ξ	Необходимая вычислительная мощность в GFLOPS								
0,1	0,014	0,022	0,029	0,061	0,092	0,123	0,238	0,356	0,475
0,01	0,023	0,034	0,045	0,086	0,129	0,172	0,319	0,479	0,639
0,001	0,027	0,040	0,053	0,102	0,154	0,205	0,385	0,578	0,770
0,0001	0,031	0,046	0,061	0,119	0,178	0,238	0,451	0,676	0,901
K	128	128	128	256	256	256	512	512	512
$N=f_q$	8	12	16	8	12	16	8	12	16

Измерение K – это количество частотных каналов в спектрограмме, оно зависит от приложения. Измерение L маскирующей матрицы H зависит от требуемой точности вычислений и используемой шкалы частот f_k . Представленные значения вычислительной сложности приведены в таблице 1 позволяют сделать вывод о необходимости создания новых алгоритмов, оптимизирующих вычислительную сложность «наивного» алгоритма одновременной маскировки. Необходимо отметить, что для повышения точности за счет увеличения числа каналов вычислительная сложность растет нелинейно.

2. Вычисление одновременной маскировки путем цифровой фильтрации. В общем случае формула одновременной маскировки не сводится к цифровой фильтрации, и оптимальный алгоритм вычисления ненамного превосходит наивный. Однако если ввести ограничения на маскирующую матрицу:

$$\forall k_1, k_2, l: H_{k_1 l} = H_{k_2 l}, \quad (5)$$

то матрица H_{KL} может быть заменена любой своей строкой H_L . С учетом особенностей вычисления k' можно заменить $S_{k'}$ на расширенный вектор S'_{k+L} :

$$\forall k = 0 \dots K - 1, l = L - 1: S'_{k+l} = S_k. \quad (6)$$

Тогда формула (1) для вычисления одновременной маскировки принимает вид:

$$M_k = \sum_{l=0}^{L-1} |S'_{k+l}|^2 \cdot H_l > |S_k|^2 = (|S'_{k+l}|^2 \otimes H_L)_{k+L/2} > |S_k|^2, \quad (7)$$

где \otimes – оператор линейной свертки. Поскольку существуют эффективные методы вычисления линейной свертки, то введенное ограничение (7) приводит к снижению требуемой вычислительной мощности алгоритма. Для выяснения условий, при которых ограничение (7) будет выполняться, необходимо описать вычисление матрицы H_{KL} . Матрица H_{KL} вычисляется из маскирующей функции $H(x, y)$, связанной со шкалой частот $f(x)$ следующими соотношениями:

$$\forall x, y \in \mathbb{R}: \\ H(x, y) = \exp\left(-\frac{(f(x+y) - f(x))^2}{2\sigma(x)^2}\right), \quad (8)$$

$$\sigma(x) = \frac{f(x)}{2.4 \cdot Q(x) \cdot \delta}, \quad (9)$$

$$Q(x) = B \cdot \frac{f(x)}{\alpha \cdot f(x) + \beta}, \quad (10)$$

$$\alpha, \beta, \delta, B = \text{const.} \quad (11)$$

Элементы матрицы H_{KL} вычисляются следующим образом:

$$\forall k = 0 \dots K - 1, l = 0 \dots L_{max} - L_{min}:$$

$$H_{kl} = H(k, k + L_{min}) / \sum_{i=L_{min}}^{L_{max}} H(k, i).$$

Значения L_{min} и L_{max} вычисляются исходя из требуемой точности ξ . Исходя из ограничения (5) произведем подстановку:

$$\forall k_1, k_2, l \in \mathbb{Z}: H_{k_1 l} = H_{k_2 l}, \\ \forall x_1, x_2, y \in \mathbb{R}: H(x_1, y) = H(x_2, y), \\ \exp\left(-\frac{(f(x_1+y) - f(x_1))^2}{2\sigma(x_1)^2}\right) = \exp\left(-\frac{(f(x_2+y) - f(x_2))^2}{2\sigma(x_2)^2}\right), \\ \frac{f(x_1+y) - f(x_1)}{\alpha \cdot f(x_1) + \beta} = \frac{f(x_2+y) - f(x_2)}{\alpha \cdot f(x_2) + \beta}.$$

При $\alpha \neq 0$ справедливо:

$$\frac{\alpha \cdot f(x_1+y) + \beta}{\alpha \cdot f(x_1) + \beta} = \frac{\alpha \cdot f(x_2+y) + \beta}{\alpha \cdot f(x_2) + \beta}.$$

Введем следующие обозначения:

$$\gamma = \frac{\beta}{\alpha}, \quad g(x) = f(x) + \gamma. \quad (12)$$

Получаем:

$$\begin{aligned} \frac{g(x_1 + y)}{g(x_1)} &= \frac{g(x_2 + y)}{g(x_2)}, \\ \frac{g(x + y)}{g(x)} &= h(y). \end{aligned} \quad (13)$$

Условие (13) накладывает ограничение на используемую шкалу частот. Очевидно, что степенная функция $g(x) = \exp(ax + b)$ удовлетворяет этому условию. Используя $g_1(x) = \ln g(x)$, $h_1(x) = \ln h(x)$, $g_2(x) = g_1(x) - g_1(0)$ получаем:

$$\begin{aligned} \frac{g(x + y)}{g(x)} &= h(y), \\ g_1(x + y) - g_1(x) &= h_1(y), \\ g_1(0 + y) - g_1(0) &= h_1(y), \\ g_1(x + y) - g_1(x) &= g_1(y) - g_1(0), \\ g_2(x + y) - g_2(x) &= g_2(y), \\ g_2(x + y) &= g_2(x) + g_2(y). \end{aligned} \quad (14)$$

Условие (14) – это функциональное уравнение Коши, которое имеет только одно непрерывное вещественное решение $g_2(x) = ax$, где $a \in \mathbb{R}$ [12]. Следовательно, $f(x)$ имеет вид:

$$\begin{aligned} f(x) = g(x) - \gamma &= \exp(g_1(x)) - \gamma = \exp(g_2(x) + g_1(0)) - \gamma, \\ f(x) &= \exp(ax + b) - \gamma, \quad a, b, \gamma = \text{const}. \end{aligned} \quad (15)$$

При выполнении условия (15) имеется возможность вычисления одновременной маскировки через цифровую фильтрацию.

Оптимальным вариантом для цифровой фильтрации для больших значений длины окна L является использование теоремы о свертке и быстрого преобразования Фурье. В этом случае может быть использован метод перекрытия с накоплением [13,14]. Выведем формулу для количества операций, необходимых алгоритму, основанному на цифровой фильтрации, для обработки входного массива, соответствующего одной секунде. Пусть входной сигнал имеет частоту дискретизации f_q , спектрограмма имеет K каналов, а окно фильтрации

для расчета одновременной маскировки имеет размер L . Поскольку на каналах с номерами $k < L$ и $k \geq K - L$ необходимо «расширение» области частот (см. также формулу (6)), то к каждому входному вектору M_K необходимо добавить по $L/2$ отсчетов с каждой стороны (см. рисунок 2). Тогда входная матрица размера $f_q \times K$ для фильтрации преобразуется в вектор длиной $f_q \times (K + L)$.

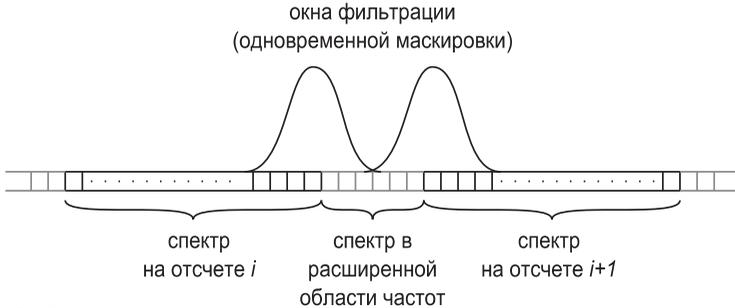


Рис. 2. Иллюстрация расширения области частот спектра для расчета одновременной маскировки

Пусть далее используется алгоритм расчета циклической свертки размера N , который требует $M(N)$ умножений, $A(N)$ сложений, $S(N) = M(N) + A(N)$. Тогда для обработки одной секунды входных данных, необходимо P операций, где:

$$P(K, L, N, f_q) = S(N) \cdot f_q \cdot (K + L) / (N - L). \quad (16)$$

Член $S(N)$ в формуле (16) зависит от алгоритма вычисления циклической свертки размера N . Сравним алгоритмы с точки зрения минимизации количества операций.

При использовании алгоритма циклической свертки на основе БПФ вычисления ведутся фактически с комплексными числами. Поэтому данный алгоритм можно использовать для вычисления двух вещественных циклических свертки одновременно. Это обстоятельство следует учесть при расчете вычислительной сложности алгоритма на основе БПФ. Формулы (17) содержат описание данного алгоритма:

$$\begin{aligned} s_1 &= g \otimes d_1 \quad s_2 = g \otimes d_2 \quad G = \Phi(g), \\ d &= d_1 + i \cdot d_2 \quad D = \Phi(d), \\ S &= G \times D \quad s = \widehat{\Phi}(S), \\ s_1 &= \Re_s \quad s_2 = \Im_s. \end{aligned} \quad (17)$$

В формулах (17) оператор \otimes означает циклическую свертку, оператор \times – поточечное умножение векторов, операторы Φ и $\hat{\Phi}$ – прямое и обратное преобразование Фурье.

В таблице 2 приведен расчет вычислительной мощности в GFLOPS, необходимой для вычисления одновременной маскировки в реальном времени, для двух различных алгоритмов циклической свертки различной длины: итерационного алгоритма Тоома-Кука для $n=4$ и алгоритма на основе БПФ, а также для наивного алгоритма.

Таблица 2. Расчеты требуемой мощности вычислительной машины для расчета одновременной маскировки в реальном времени

Размер (N)	Необходимая вычислительная мощность в GFLOPS								
	Наивный алгоритм								
—	0,014	0,034	0,053	0,061	0,129	0,205	0,238	0,479	0,770
Алгоритм циклической свертки на основе БПФ Кули-Тьюки									
16	0,027	0,075	0,168	0,486					
32	0,026	0,047	0,070	0,075	0,179	0,380	0,854		
64	0,028	0,047	0,066	0,066	0,115	0,172	0,184	0,394	0,783
128	0,032	0,051	0,071	0,069	0,112	0,157	0,157	0,267	0,398
256	0,037	0,058	0,079	0,076	0,120	0,165	0,161	0,258	0,362
512	0,042	0,065	0,088	0,085	0,132	0,180	0,174	0,272	0,375
Итерационный алгоритм циклической свертки Тоома-Кука по основанию 4									
16	0,039	0,108	0,244	0,702					
32	0,057	0,104	0,155	0,167	0,395	0,840	1,887		
64	0,064	0,107	0,150	0,150	0,262	0,391	0,420	0,898	1,786
128	0,100	0,160	0,220	0,215	0,349	0,490	0,491	0,834	1,239
256	0,124	0,194	0,265	0,257	0,404	0,556	0,545	0,870	1,222
512	0,186	0,290	0,393	0,379	0,589	0,803	0,779	1,216	1,673
K	128	128	128	256	256	256	512	512	512
f_q (кГц)	8	12	16	8	12	16	8	12	16
ξ	0,1	0,01	0,001	0,1	0,01	0,001	0,1	0,01	0,001
L	7	11	13	15	21	25	29	39	47

Из таблицы 2 видно, что алгоритм на основе БПФ требует меньше вычислений, чем итерационный алгоритм. В некоторых случаях – для малой длины L – наивный алгоритм оказывается лучше алгоритма БПФ. Однако, при необходимости вычислений с высокой точностью (ξ) или разрешающей способностью (K) значение L будет достаточно большим, и использование фильтрации позволит сократить вычислительную сложность до 2 раз.

В разработанном алгоритме (см. рисунок 3) используются специальные объекты входного потока $istr$ и выходного потока $ostr$, которые инкапсулируют в себе функционал вычисления расширенных входных

векторов по формуле 6 перед фильтрацией и последующего «сужения» после фильтрации.

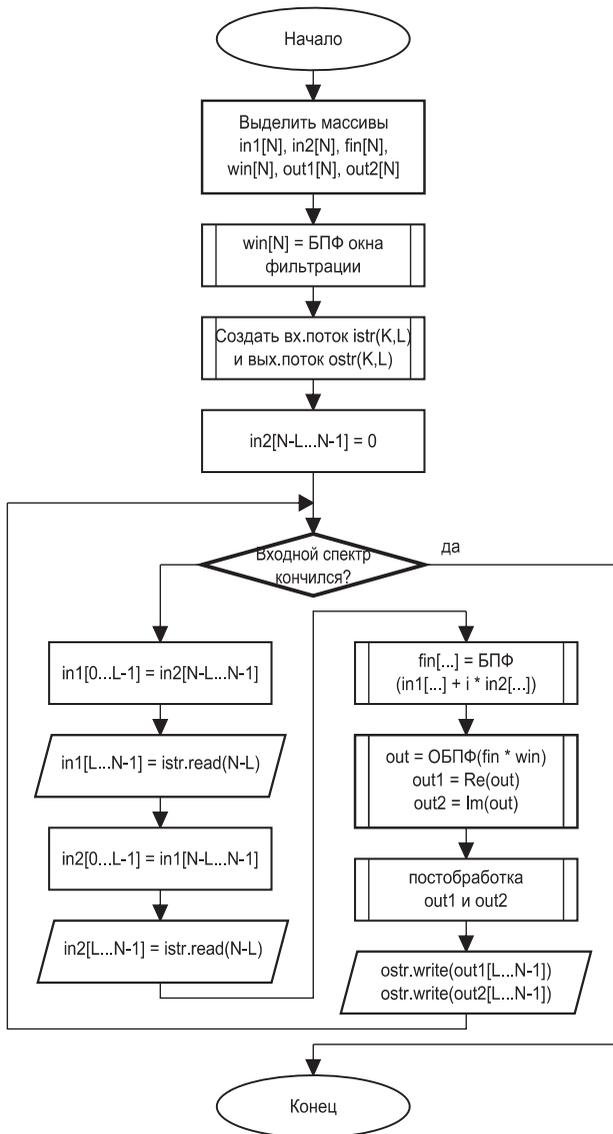


Рис. 3. Алгоритм быстрого вычисления одновременной маскировки на основе БПФ

Использование таких потоков позволяет сделать независимыми границы входных блоков, соответствующих временным отсчетам, (размера K) и блоков для фильтрации (размера M). В алгоритме за одну итерацию вычисляется два выходных блока, т.е. $2(N - L)$ чисел. При этом, поскольку используется метод пересечения с накоплением, входные блоки пересекаются – поэтому используется копирование из одной части входного блока в другую. Постобработка выходных векторов (предпоследний шаг в итерации) заключается в выполнении операции сравнения (оператор $>$) из формулы (7).

3. Заключение. Как следует из таблицы 2, предложенный метод позволяет вычислять одновременную маскировку быстрее, чем оригинальный (наивный) алгоритм при определенных условиях. Эти условия заключаются в выполнении ограничения (15) на используемую шкалу частот и достаточно больших значениях K и ξ (см. таблицу 2). Требуемая вычислительная мощность при этом снижается до 2 раз.

Значения вычислительной сложности как для оригинального (до 0,77 GFLOPS), так и для предлагаемого (до 0,36 GFLOPS) алгоритмов позволяют вычислять одновременную маскировку в реальном времени с учетом вычислительной мощности современных цифровых процессоров (от 3 GFLOPS до более чем 100 GFLOPS). Однако вычисление одновременной маскировки, как правило, не является конечной целью в прикладных задачах, а является лишь промежуточной ресурсоемкой задачей. Поэтому снижение сложности вычисления одновременной маскировки позволяет повысить шансы на работу всей системы в реальном времени, а также снизить требования к ее аппаратному обеспечению.

Литература

1. *Сапожков М.А.* Речевой сигнала в кибернетике и связи // М.: Государственное издательство литературы по вопросам связи и радио. 1963. 452 с.
2. *Бондаренко В.П., Коубинский В.П., Мецержаков Р.В.* Нестационарные модели в обработке речевых сигналов // Акустика речи. Медицинская и биологическая акустика. Архитектурная и строительная акустика и вибрации. Сб. трудов XVIII сессии Российского акустического общества. Т.3. М.: ГЕОС, 2006. С. 8-11.
3. *Ronzhin A.L., Budkov V.Yu.* Multimodal Interaction with Intelligent Meeting Room Facilities from Inside and Outside // Springer-Verlag Berlin Heidelberg, S. Balandin et al. (Eds.): NEW2AN/IUSMART 2009, LNCS 5764, 2009, pp. 77–88.
4. *Yusupov R.M., Ronzhin An.L., Prischepa M., Ronzhin A.L.* Models and Hardware-Software Solutions for Automatic Control of Intelligent Hall // Automation and Remote Control. 2011. vol. 72. no. 7. pp. 1389–1397.
5. *Ronzhin A., Karpov A., Kipyatkova I., Zelezny M.* Client and Speech Detection System for Intelligent Infokiosk // Springer-Verlag Berlin Heidelberg, Petr Sojka et al. (Eds.): TSD 2010, LNAI 6231, 2010. pp. 560–567.
6. *Meshcheryakov R.V., Bondarenko V.P.* Dialogue as a basis for construction of speech systems // Cybernetics and Systems Analysis. 2008. no. 44(2). pp. 175-184.
7. *Meshcheryakov R. V., Balatskaya L. N., Choinzonov E. L., Chizevskaya S. Yu., Kostyuchenko E.U.* Software for Assessing Voice Quality in Rehabilitation of Patients after Surgical Treatment of Cancer of Oral Cavity, Oropharynx and Upper Jaw // Proceedings of the 15th International Conference, SPECOM 2013, September 1-5, 2013, Pilsen, Czech Republic, 2013. pp 294-301.

8. Bondarenko V.P., Kvasov A.N., Konev A.A., Meshcheryakov R.V., Choinzonov E.L., Chizhevskaya S.Y. Software for speech signal research in patients with malignant diseases of the throat // *Biomedical Engineering*, 2009. no. 43(4). pp. 183-186.
9. Ponizov A.G., Meshcheryakov R.V. Implementation of Tuning Fork Hearing Tests Using Electronic Devices // *Biomedical Engineering*. 2012. no. 46(1). pp. 32-35.
10. Меццержаков Р.В. Структура систем синтеза и распознавания речи // *Известия Томского политехнического университета*. 2009. Т. 315. № 5. С. 121-126.
11. Bondarenko V.P., Ponomarev A.A., Rogozinskaya E.A. Модель одновременной маскировки // *Интеллектуальные системы в управлении, конструировании и образовании*. Томск: STT, 2004. С. 167-174.
12. Huang X., Acero A., Hon H.-W. Spoken Language Processing: a guide to theory, algorithm, and system development // Prentice Hall, 2001. 980 p.
13. Jung S. Hyers-Ulam-Rassias Stability of Functional Equations in Nonlinear Analysis // Springer optimization and its applications. Springer, 2011. URL: <http://books.google.ru/books?id=NcClS-Hm19kC>.
14. Traunmüller H. Analytical expressions for the tonotopic sensory scale // *The Journal of the Acoustical Society of America*. 1990. vol. 88, no. 1. pp. 97-100. URL: <http://scitation.aip.org/content/asa/journal/jasa/88/1/10.1121/1.399849>.

References

1. Sapozkov M.A. *Rechevoj signala v kibernetike i svyazi* [Speech signal in cybernetics and communication]. M., Gosudarstvennoe izdatelstvo literatury po voprosam svyazi I radio, 1963. 452 p. (In Russ.).
2. Bondarenko V.P., Kotsubinsky V.P., Meshcheryakov R.V. [Nestacionarnie modeli v obrabotke rechevix signalov]. *Acistica rechi. Medicinskaja i biologicheskaja akustika. Arhitekturnaja i stroitel'naja akustika i vibracii. Sbornik trudov XVIII sessii Russain acoustic society* [Acoustics of speech. Medical and biological acoustics. Architectural and building acoustics and vibration. Proceedings of the 17th Session of the Russian Acoustical Society]. V.3, M: GEOS, 2006. pp. 8-11. (In Russ.).
3. Ronzhin A.L., Budkov V.Yu. Multimodal Interaction with Intelligent Meeting Room Facilities from Inside and Outside // Springer-Verlag Berlin Heidelberg, S. Balandin et al. (Eds.): NEW2AN/ruSMART 2009, LNCS 5764, 2009, pp. 77–88.
4. Yusupov R.M., Ronzhin An.L., Prischepa M., Ronzhin A.L.L. Models and Hardware-Software Solutions for Automatic Control of Intelligent Hall. *Automation and Remote Control*. 2011. vol. 72, no. 7. pp. 1389–1397.
5. Ronzhin A., Karpov A., Kipyatkova I., Zelezny M. Client and Speech Detection System for Intelligent Infokiosk. Springer-Verlag Berlin Heidelberg, Petr Sojka et al. (Eds.): TSD 2010, LNAI 6231. 2010. pp. 560–567.
6. Meshcheryakov R.V., Bondarenko V.P. Dialogue as a basis for construction of speech systems. *Cybernetics and Systems Analysis*, 2008. no. 44 (2). pp. 175-184.
7. Meshcheryakov R. V., Balatskaya L. N., Choinzonov E.L., Chizevskaya S.Yu., Kostyuchenko E.U. Software for Assessing Voice Quality in Rehabilitation of Patients after Surgical Treatment of Cancer of Oral Cavity, Oropharynx and Upper Jaw. Proceedings of the 15th International Conference, SPECOM 2013, September 1-5, 2013, Pilsen, Czech Republic, 2013. pp. 294-301.
8. Bondarenko V.P., Kvasov A.N., Konev A.A., Meshcheryakov R.V., Choinzonov E.L., Chizhevskaya S.Y. Software for speech signal research in patients with malignant diseases of the throat. *Biomedical Engineering*, 2009. no. 43 (4). pp. 183-186.
9. Ponizov A.G., Meshcheryakov R.V. Implementation of Tuning Fork Hearing Tests Using Electronic Devices. *Biomedical Engineering*, 2012. no. 46 (1), pp. 32-35.
10. Meshcheryakov R.V. [Structure of speech synthesis and recognition systems]. *Isvestiya Tomskogo politehnicheskogo universiteta – Tidings Tomsk Polytechnic University*. 2009. vol. 315. no. 5. pp 121-126. (In Russ.).
11. Bondarenko V.P., Ponomarev A.A., Rogozinskaya S.I. [Model odnovenmennoy maskirovki]. *Intellektualniye sistemi v upravlenii, konstruirovanii i obrazovanii – Intelligent systems in the management, engineering and education*. Tomsk, STT. 2004. pp. 167-174. (In Russ.).

12. Huang X., Acero A., Hon H.-W. Spoken Language Processing: a guide to theory, algorithm, and system development. Prentice Hall, 2001. 980 p.
13. Jung S. Hyers-Ulam-Rassias Stability of Functional Equations in Nonlinear Analysis. Springer optimization and its applications. Springer, 2011. Available at: <http://books.google.ru/books?id=NcClS-Hm19kC>.
14. Traummuller H. Analytical expressions for the tonotopic sensory scale. The Journal of the Acoustical Society of America. 1990. vol. 88, no. 1. pp. 97-100. Available at: <http://scitation.aip.org/content/asa/journal/jasa/88/1/10.1121/1.399849>.

Тиунов Сергей Дмитриевич — аспирант кафедры Комплексной информационной безопасности электронно-вычислительных систем ТУСУР. Область научных интересов: обработка речевого сигнала, математическое моделирование. Число научных публикаций — 17. t5d@mail.ru; г. Томск, пр. Ленина, 40, ауд. 210, тел. +7 (913) 823 32 56.

Tiunov Sergey Dmitrievich — post-graduate student, Dept. of Complex Security of Electronic-computing Systems of Tomsk State University of Control Systems and Radioelectronics (TUSUR). Research interests: speech signal processing, mathematical modeling. The number of publications — 17. t5d@mail.ru; KIBEVS Dept. TUSUR, 40, Lenin-avenue Tomsk, 634050, Russia; phone +7 (913) 832 32 56.

Мещеряков Роман Валерьевич — д-р техн. наук, профессор кафедры Комплексной информационной безопасности электронно-вычислительных систем ТУСУР. Область научных интересов: системный анализ, анализ и синтез речи информационная безопасность, вопросы обработки информации в интеллектуальных системах, особое внимание уделяется вопросам создания информационно-безопасных систем. Число научных публикаций — 247. mrv@security.tomsk.ru; 634050, г. Томск, пр. Ленина, 40, ауд. 210; п.т. +7 (3822) 900-111, факс +7 (3822) 900-111.

Meshcheryakov Roman Valerievich — Ph.D., Dr. Sci., professor, Dept. of Complex Security of Electronic-computing Systems of Tomsk State University of Control Systems and Radioelectronics (TUSUR). Research interests: speech analysis, speech recognition, medical technology, information security. The number of publications — 247, IEEE Senior Member. mrv@security.tomsk.ru; KIBEVS Dept. TUSUR, 40, Lenin-avenue Tomsk, 634050, Russia; office phone +7(3822)413-426, fax +7(3822)900-111.

Черных Дмитрий Владимирович — аспирант кафедры Комплексной информационной безопасности электронно-вычислительных систем ТУСУР. Область научных интересов: обработка речевого сигнала, математическое моделирование. Число научных публикаций — 9. dmit-cher@mail.ru; г. Томск, пр. Ленина, 40, ауд. 210, тел. +7 (952) 805 97 59.

Chernyh Dmitriy Vladimirovich — post-graduate student, Dept. of Complex Security of Electronic-computing Systems of Tomsk State University of Control Systems and Radioelectronics (TUSUR). Research interests: speech signal processing, mathematical modeling. The number of publications — 9. dmit-cher@mail.ru; KIBEVS Dept. TUSUR, 40, Lenin-avenue Tomsk, 634050, Russia; phone +7 (952) 805 97 59.

Поддержка исследований. Работа выполнена при финансовой поддержке РФФИ (проект № 14-07-00449 -а), государственного задания Министерства образования и науки РФ, проект № 1220 «Фундаментальные основы проектирования информационно-безопасных систем».

Acknowledgements. This research is supported by RFBR (grant 14-07-00449 -а) and the state contract of the Ministry of Education and Science №1220 “Fundamentals of information-security systems designing”.

РЕФЕРАТ

Тиунов С.Д., Мещеряков Р.В., Черных Д.В. **Оптимизация вычисления одновременной маскировки речевого сигнала.**

При обработке речи проводятся различные операции над звуковым сигналом и, как правило, это требует значительных вычислительных ресурсов и не может быть проведено в реальном масштабе времени. Это связано с одной стороны с тем, что при разработке новых алгоритмов для обработки речи на первом месте стоит вопрос точности, а не вопрос вычислительной эффективности. Вместе с тем некоторые приложения речевых технологий подразумевают обработку сигнала в реальном времени. Предлагается эффективный численный метод для вычисления одновременной маскировки из модели слуховой системы, что позволяет применять ее в программном комплексе для анализа речевого сигнала в норме и при патологии.

Одновременная маскировка – это особенность восприятия звука слуховой системой человека. Эта особенность заключается в том, что из двух близких частотных составляющих ухом будет воспринята только та, которая имеет большую интенсивность. Моделирование эффекта одновременной маскировки позволяет выделить в спектрограмме сигнала наиболее важные (информативные) составляющие с точки зрения слуховой системы человека.

В работе вводится ограничение на используемую шкалу частот, что позволяет свести вычисление одновременной маскировки к вычислению линейной свертки. Поскольку существуют эффективные методы вычисления линейной свертки, то введенное ограничение приводит к снижению вычислительной сложности алгоритма.

Предложенный метод позволяет вычислять одновременную маскировку быстрее, чем оригинальный (наивный) алгоритм при определенных условиях. Эти условия заключаются в выполнении ограничения на используемую шкалу частот и определенные (высокие) значения разрешающей способности. Требуемая вычислительная мощность при этом снижается до 2 раз.

Значения вычислительной сложности как для оригинального (до 0,77 GFLOPS), так и для предлагаемого (до 0,36 GFLOPS) алгоритмов позволяют вычислять одновременную маскировку в реальном времени с учетом вычислительной мощности современных цифровых процессоров (от 3 GFLOPS до более чем 100 GFLOPS). Однако вычисление одновременной маскировки, как правило, не является конечной целью в прикладных задачах, а является лишь промежуточной ресурсоемкой задачей. Поэтому снижение сложности вычисления одновременной маскировки позволяет повысить шансы на работу всей системы в реальном времени, а также снизить требования к ее аппаратному обеспечению.

SUMMARY

Tiunov S.D., Meshcheryakov R.V., Chernykh D.V. **Optimization of calculation of frequency masking of speech signal.**

Speech processing is associated with various operations on sound signal and typically requires significant computational resources and may not be conducted in real time. On the one hand this is due to the fact that the development of new algorithms for speech processing in the first place is a question of accuracy, not a question of computational efficiency. However, some applications involve speech signal processing techniques in a real time. An efficient numerical method for calculating the frequency masking of the auditory system model is proposed, which allows using it in a software package for the analysis of speech signal in normal and pathological conditions.

Frequency masking is a feature of the perception of sound in human auditory system. It states that of the two neighboring frequency components only one that has a greater intensity will be heard. Simulation of frequency masking allows selecting the most important (informative ones) components from the spectrogram of the signal from the perspective of human auditory system.

The constraint on the used frequency scale is proposed and considered, that allows calculating frequency masking by calculating linear convolution. Since the efficient methods of convolution calculation are known, the proposed constraint leads to reduce the computational complexity of the algorithm.

The proposed method allows calculating the frequency masking faster than the original (naive) algorithm under certain conditions. These conditions include the constraint on the used frequency scale and certain (high) resolution parameters. The computational power required in this case is reduced up to 2 times.

Computational complexity for the original (up to 0,77 GFLOPS), and for the proposed (up to 0,36 GFLOPS) algorithms allows us to calculate the frequency masking in real time based on computational power of modern CPUs (3 GFLOPS to more than 100 GFLOPS). However, the calculation of frequency masking is usually not an end goal in applications, but only an intermediate resource-consuming task. Therefore, reducing the complexity of calculating the frequency masking can increase the chances of the entire system real time work, as well as reduce its hardware requirements.