

СИСТЕМИ ТА МЕТОДИ ОБРОБКИ ІНФОРМАЦІЇ

УДК 621.3

В.И. Соловьев,

кандидат технических наук, доцент

О.В. Рыбальский,

доктор технических наук, профессор

ОПТИМИЗАЦИЯ ВЫЧИСЛИТЕЛЬНОЙ СЛОЖНОСТИ СПЕКТРАЛЬНОГО АНАЛИЗА В СОВРЕМЕННЫХ ЗАДАЧАХ РЕЧЕВЫХ ТЕХНОЛОГИЙ

В статье рассмотрены результаты исследований особенностей построения быстрых вычислительных схем спектрального анализа сигналов речи на основе ортогонального и неортогонального время-частотного преобразования. Подход опирается на рекурсивные методы построения схем спектрального анализа. Приведены результаты исследований по применению полученных вычислительных процедур на основе неортогонального базиса для анализа речевых файлов. Показано, что благодаря существенному снижению вычислительной сложности построения спектров речевых фрагментов возможно проведение вычислений в режиме реального времени для большинства задач речевых технологий на современных персональных компьютерах. Полученные результаты могут быть применены при построении эффективных вычислительных схем в области речевых технологий.

Ключевые слова: частотный спектр, ортогональный базис, неортогональный базис, фонема, форманта, характеристики голоса, вычислительная сложность.

У статті розглянуто результати досліджень особливостей побудови швидких обчислювальних схем спектрального аналізу сигналів мови на основі ортогонального і неортогонального часо-частотного перетворення. Підхід спирається на рекурсивні методи побудови схем спектрального аналізу. Наведені результати досліджень із застосування отриманих обчислювальних процедур на основі неортогонального базису для аналізу мовних файлів. Показано, що завдяки істотному зниженню обчислювальної складності побудови спектрів мовних фрагментів можливе проведення обчислень у режимі реального часу для більшості завдань мовних технологій на сучасних персональних комп'ютерах. Отримані результати можуть бути застосовані при побудові ефективних обчислювальних схем в галузі мовних технологій.

Ключові слова: частотний спектр, ортогональний базис, неортогональний базис, фонема, форманта, характеристики голоса, обчислювальна складність.

The results of researches of features of construction of rapid calculable charts of spectrology of signals of speech are considered on the basis of ortogonal and unortogonal time-frequency transformation. The results of the researches are brought on an application of the got calculable

procedures on the basis of unortogonal base for the analysis of speech files. It is shown that due to the substantial decline of calculable complication of the construction of spectrums of speech fragments realization of calculations it is possible to carry out real-time for most tasks of speech technologies on the modern personal computers. The results can be applied at the construction of effective calculable charts in area of speech technologies.

Keywords: *frequency spectrum, ortogonal basis, unortogonal basis, phoneme, formant, descriptions of voice, calculable complication.*

При решении большинства современных задач речевых технологий применяются различные модификации спектрального анализа. Это связано с физической моделью акустических процессов речеобразования. Современная акустическая теория речеобразования основывается на работах Гельмгольца [1]. Эти работы, кроме фундаментальных физических концепций, по существу, определили и математические методы исследования задач речевых технологий – спектральный анализ на основе преобразования Фурье [2]. Существенная детализация и развитие исследований в классических трудах многих исследователей за прошедшие 150 лет не привнесли принципиально новых методов в спектральный анализ в области речевых технологий. Это касается прежде всего фундаментальных трудов Фанта [3] и Фланагана [4].

Вторая половина 20-го столетия существенно модифицировала подход к методам спектрального анализа. Были разработаны быстрые алгоритмы спектрального анализа, в том числе алгоритмы быстрых преобразований Фурье (БПФ) [2]. Они широко используются в современных стандартах эффективного сжатия аудиоинформации, основой которых является формат mp3. Развитие вейвлет-анализа также привело к дальнейшей эволюции быстрых вычислений [2]. Разумеется, на сегодняшний день в различных областях применения спектрального анализа быстрые алгоритмы вычислений не имеют альтернативы. Но при этом большинство исследователей применяют быстрые алгоритмы в какой-то мере автоматически, считая их оптимальными с вычислительной точки зрения. Все это в полной мере относится практически ко всем задачам речевых технологий [5–11].

Но в ряде актуальных задач речевых технологий нет возможности использовать БПФ. Например, известно, что слуховые нейроны мозга анализируют акустическую информацию на малых временных интервалах порядка 20 мс с разрешающей способностью по частоте порядка 1 Гц [5]. Следовательно, для исследования сигналов на малых временных интервалах с высоким частотным разрешением необходимо отказаться от ортогонального преобразования Фурье, что накладывает запрет на применение практически любых классических алгоритмов БПФ [2].

Кроме того, важным моментом речевых технологий является возможность построения спектра с дискретностью шага по времени равного частоте дискретизации акустической информации, что слуховой анализатор человека реализует в реальном масштабе времени. На современных персональных

компьютерах такие задачи не поддаются решению в приемлемое время даже при использовании БПФ (соответствующие оценки будут приведены далее).

В данной статье представлены результаты исследований, которые позволяют существенно снизить вычислительную сложность спектрального анализа в задачах речевых технологий как для ортогонального базиса Фурье, так и для неортогонального базиса функций.

1. Постановка задачи исследования

Приводимые далее исследования могут применяться в различных областях спектрального анализа. Однако актуальность использования подобных методов существенно зависит от области спектрального анализа. Далее все рассмотрения приводятся для задач речевых технологий.

Рассмотрим графики фрагментов речи на небольших временных интервалах и их спектр для ортогонального и неортогонального преобразования Фурье (см. рис. 1 и рис. 2 соответственно). На представленных аудиофрагментах речи частота дискретизации 8000 Гц, размеры временного окна – 20 мс. Шаг по частоте дискретизации для ортогонального преобразования – 50 Гц, для неортогонального преобразования – 1 Гц.

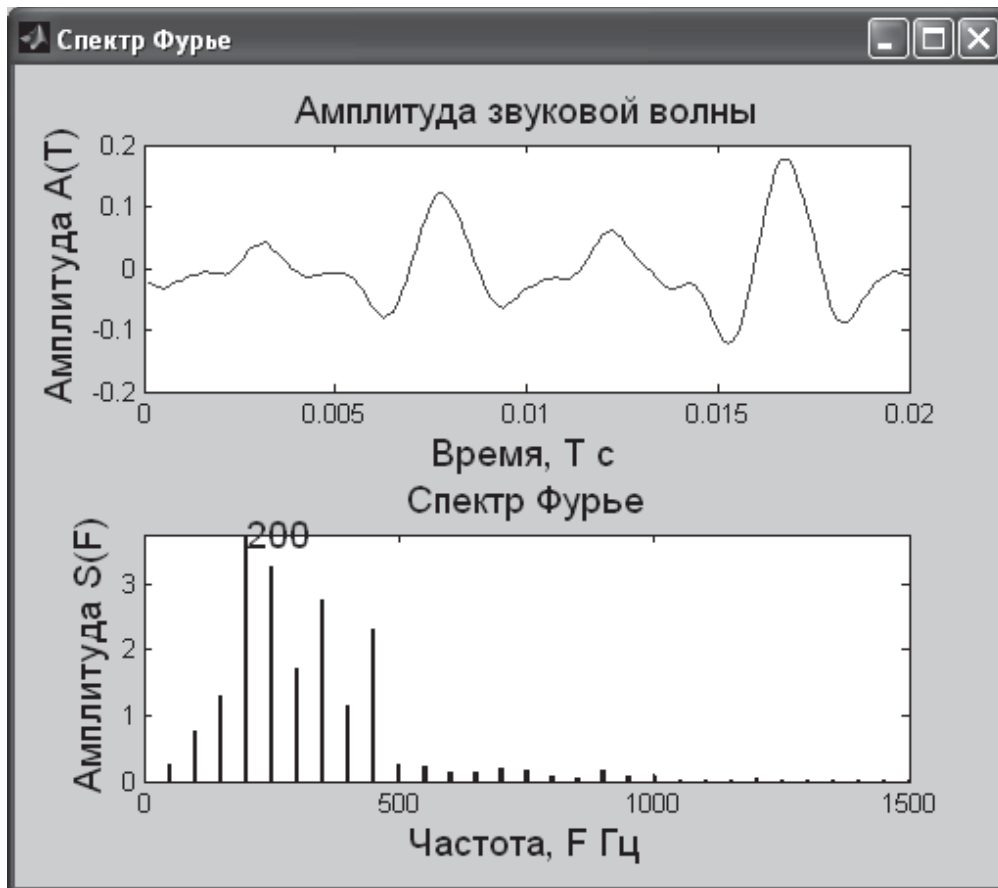


Рис. 1. Спектр Фурье с ортогональным преобразованием

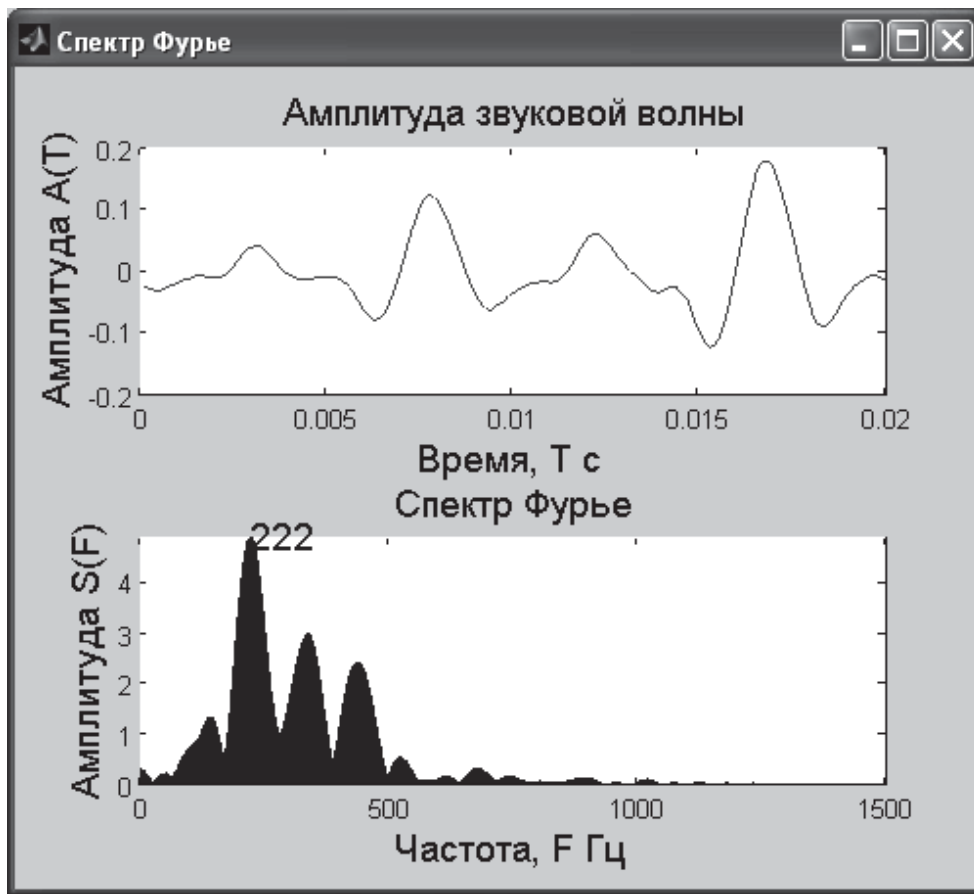


Рис. 2. Спектр Фурье с неортогональным преобразованием

Как видно из сравнения спектров, представленных на рис. 1 и рис. 2, при неортогональном преобразовании результаты анализа дают более точные результаты и по своему качеству при малом шаге частоты приближаются к непрерывному преобразованию Фурье.

В данном исследовании нас будет интересовать вопрос оптимизации вычислительной сложности преобразования Фурье. Представленные на рис. 1 и рис. 2 спектры вычислялись на основе классического преобразования Фурье [2]:

$$S(k) = \text{abs} \left(\sum_{t_i=0}^{N_m} [\exp(-2\pi j / N)(t-1)(k-1)] A(t) \right), \quad (1)$$

где $j = \sqrt{-1}$, abs – модуль комплексного преобразования Фурье, $S(k)$ – амплитуда спектра, $A(t)$ – амплитуда звуковой волны для дискретного отсчета t_i .

Параметр k в (1) будем рассматривать не как дискретную переменную номера частоты спектра, а как непрерывную переменную, которая может изменяться в диапазоне $1 \leq k \leq N/2$. Для дискретных значений k в этом диапазоне выражение (1) является обычным оконным дискретным преобразованием Фурье с ортогональным базисом. Для дробных значений k это выражение является преобразованием с неортогональным базисом и произвольной частотой – $2\pi k / N$. Теория

інформації не виключає фізичного існування будь-яких частотних складових в діапазоні частот від $1/T$ до частоти Найквіста (T – часовий інтервал перетворення).

При розрахунку спектра для N дискретних відліків амплітуди звукової хвилі на основі БПФ вимагається $2N \log_2 N$ дійсних операцій додавання і множення. При розрахунку спектра для N відліків амплітуди звукової хвилі і $M > N$ частот вимагається $M \cdot N$ дійсних операцій додавання і множення.

Розглянемо достатньо актуальну задачу розрахунку дискретного перетворення Фур'є для N дискретних відліків амплітуди звукової хвилі зі зсувом інтервалу перетворення на мінімально можливий часовий інтервал, рівний часу дискретизації відліків. Ця задача емулює з точністю до частоти дискретизації неперервні частотні перетворення нейронів слухового аналізатора людини. Нехай у нас є Q дискретних відліків амплітуди звукової хвилі на великому часовому інтервалі (більше 10 секунд). Оцінимо обчислювальну складність дискретного перетворення Фур'є для невеликого часового інтервалу порядку 100 мс (приблизня середня тривалість голосного звуку) на частоті дискретизації 41400 Гц.

Для БПФ це величина $2(Q-N) \log_2 N$ дійсних операцій додавання і множення, де $N = 4140$, $Q = 441000$ (для 10 секунд). Це величина порядку $4 \cdot 10^9$ дійсних операцій додавання і множення. Так, наприклад, обчислення на основі БПФ для всього часового інтервалу на сучасному комп'ютері з частотою процесора 2 ГГц вимагає в середньому 10^6 секунд. Естествено, ортогональний варіант перетворення практично неможливо реалізувати для великих фрагментів мовних файлів.

Сучасні засоби паралелізації обчислень, наприклад на графічних адаптерах, допомагають виконувати розрахунки за прийнятний час.

Поставимо задачу створення більш ефективного, з точки зору обчислювальної складності, алгоритму розрахунку спектрів подібних задач в мовних технологіях.

2. Рекурсивна схема розрахунку спектрів Фур'є

Для подальших висновків введемо позначення:

$$MC(t, k) = \begin{pmatrix} \cos[(-2\pi/N)(t-1)(k_1-1)] & \dots & \cos[(-2\pi/N)(t-1)(k_1-1)] \\ \dots & \dots & \dots \\ \cos[(-2\pi/N)(t-1)(k_q-1)] & \dots & \cos[(-2\pi/N)(t-1)(k_q-1)] \end{pmatrix}, \quad (2)$$

$$MS(t, k) = \begin{pmatrix} \sin[(-2\pi/N)(t-1)(k_1-1)] & \dots & \sin[(-2\pi/N)(t-1)(k_1-1)] \\ \dots & \dots & \dots \\ \sin[(-2\pi/N)(t-1)(k_q-1)] & \dots & \sin[(-2\pi/N)(t-1)(k_q-1)] \end{pmatrix}, \quad (3)$$

де $MC(t, k)$, $MS(t, k)$ – частотно-часові косинус і синус матриці відповідно, N – кількість часових відліків амплітуди звукової хвилі в часовому інтервалі перетворення тривалістю T , t , k – індекси часу і частоти відповідно.

При этом $1 \leq t \leq N$ с шагом дискретности 1, $1/T \leq k \leq N/2$ с произвольным шагом дискретизации, q – число частот с выбранным шагом по частоте.

Пусть

$$\operatorname{Re} S_p = MC \cdot A_p, \quad (4)$$

$$\operatorname{Im} S_p = MS \cdot A_p, \quad (5)$$

где $\operatorname{Re} S_p$ – вещественная часть (вектор-столбец, зависящий от частоты) комплексного преобразования Фурье для временного окна в начальный дискретный момент времени p , $\operatorname{Im} S_p$ – мнимая часть (вектор-столбец, зависящий от частоты) комплексного преобразования Фурье для временного окна в начальный дискретный момент времени p , A_p – вектор столбец амплитуд звуковой волны для временного окна длительностью T , в начальный дискретный момент времени p .

При принятых обозначениях, с учетом известных формул для суммы косинуса и синуса двух углов, опуская промежуточные преобразования, можно получить рекурсивные формулы в матричной форме для расчета последовательных во времени значений $\operatorname{Re} S_p$ и $\operatorname{Im} S_p$:

$$\begin{aligned} \operatorname{Re} S_{p+1} = & \operatorname{Re} S_p \cdot MC(2, k)' - \operatorname{Im} S_p \cdot MS(2, k)' + \\ & + MC(2, k) \cdot (A(t_{p+N}) - A(t_{p-1})), \end{aligned} \quad (6)$$

$$\begin{aligned} \operatorname{Im} S_{p+1} = & \operatorname{Im} S_p \cdot MC(2, k)' + \operatorname{Re} S_p \cdot MS(2, k)' + \\ & + MS(2, k) \cdot (A(t_{p+N}) + A(t_{p-1})), \end{aligned} \quad (7)$$

где $MC(2, k)', MS(2, k)'$ – транспонированные вторые по времени вектор-столбцы матриц (2) и (3), $A(t_{p+N}), A(t_{p-1})$ – значение амплитуды звуковой волны в моменты времени t_{p+N} и t_{p-1} .

Рекурсивные формулы (6) и (7) позволяют существенно снизить вычислительную сложность рассматриваемых задач вычисления спектров. В частности, снизить число вещественных операций для условий, приведенных в постановочной части задачи, где величина порядка $6(Q - N)N$ операций сложения и умножения. Для временных окон длительностью свыше 100 мс рекурсивные процедуры на порядок быстрее. При этом важно то, что они применимы как для ортогонального, так и для неортогонального базиса.

На практике время выполнения операций, связанное с вычислительной сложностью, определяется не только числом арифметических операций, но и организацией вычислений в конкретной программной среде. Так, реализация приведенной рекурсивной схемы в системе Matlab за счет эффективной системы матричных и векторных операций позволяет снизить время расчета спектров примерно в 50–70 раз.

Выводы

Предложена рекурсивная схема вычислений спектров речевых фрагментов. Данная схема позволяет существенно снизить вычислительные затраты в ряде задач речевых технологий.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *Helmholtz H. von.* Die Lehe von Tonempfindungen. Brannschweig, Vieweg, 1863.
2. *Малла С.* Вейвлеты в обработке сигналов / С. Малла. – М. : Мир, 2005. – 670 с.
3. *Фланаган Дж.* Анализ, синтез и восприятие речи / Дж. Фланаган : Пер. с англ. / Под ред. А.А. Пирогова. – М. : Связь, 1968. – 396 с.
4. *Фант Гуннар.* Анализ и синтез речи. Пер. с англ. / Фант Гуннар ; под ред. В.С. Лозовского, Н.В. Бахмутовой, Н.Г. Загоруйко. – Новосибирск : Наука, Сиб. отд., 1970. – 167 с.
5. Психоакустические аспекты восприятия речи. Механизмы деятельности мозга / Под ред. Н.П. Бехтеревой. – М. : Наука, 1988. – 504 с.
6. *F. Elinek.* Распознавание непрерывной речи статистическими методами / F. Elinek // ТИИЭР 64. – 1976. – № 4. – С. 131–160.
7. *F. Elinek.* Разработка экспериментального устройства, распознающего отдельно произнесенные слова / F. Elinek // ТИИЭР 73. – 1985. – № 11. – С. 91–99.
8. *Цвикер Э.* Ухо как приемник информации / Э. Цвикер, Р. Фельдкеллер : Пер. с нем. / Под ред. Б.Г. Белкина. – М. : Связь, 1971. – 225 с.
9. *Алдошина И.А.* Основы психоакустики / И.А. Алдошина // Звукорежиссер. – 2000. – № 6. – С. 36–40.
10. *Сорокин В.Н.* Теория речеобразования / В.Н. Сорокин. – М. : Радио и связь, 1985. – 312 с.
11. *Рыбальский О.В.* Система идентификации аппаратуры аудиозаписи на основе мультифрактального подхода / О.В. Рыбальский, В.И. Соловьев // Вісник Східноукраїнського національного університету. – 2010. – № 9 (151). – С. 58–64.

Отримано 15.09.2014.