OSTIS-2013
(Open Semantic Technologies for Intelligent Systems)

УДК 004.942; 781.1

ОПТИМИЗАЦИЯ ОКОННОГО ФУРЬЕ И НЕПРЕРЫВНОГО ВЕЙВЛЕТ ПРЕОБРАЗОВАНИЙ ДЛЯ СПЕКТРАЛЬНОГО АНАЛИЗА МУЗЫКАЛЬНЫХ СИГНАЛОВ

Алиев Р.М.

Белорусский государственный университет,
г. Минск, Республика Беларусь

RomanAliyev@gmail.com

Предлагается метод для улучшения частотно-временного разрешения при спектральном анализе музыкальных сигналов. Этот метод позволяет найти оптимальные параметры спектральных преобразований и состоит из трех этапов – спектральное моделирование музыкальных звуков, синтез музыкальных спектров и сравнение вычисленного спектра с синтезированным спектром. В ходе экспериментов с базой данных классической музыки были исследованы оптимальный размер окна в оконном преобразовании Фурье и оптимальная компактность материнского вейвлета в непрерывном вейвлет преобразовании.

Ключевые слова: распознавание музыки; спектральный анализ; оконное преобразование Фурье; непрерывное вейвлет преобразование.

Введение

За последние тридцать лет благодаря развитию компьютерных технологий и цифровой обработки сигнала в решении задач распознавании музыки было достигнуто множество значимых результатов и было предложено множество алгоритмов [Gómez et al., 2003] [Peeters, 2004], позволяющих более качественно выделять спектральные характеристики музыкального звука. В частности, один из них – алгоритм непрерывного вейвлет преобразования получил широкий интерес в распознавании музыки после публикации работы Ж. Морле о приложении вейвлетов к анализу звуковых паттернов [Kronald-Martinet et al., 1987] и позиционируется сейчас как хорошая альтернатива Фурье преобразованию. Он является аналогом многомасштабного преобразования [Brown, 1991], т.е. позволяет выделять низкие и высокие частоты сигнала в разных масштабах времени. Потому, безусловно, в отличие от Фурье преобразования, вейвлет-анализ предоставляет возможности для получения спектра звука с лучшей частотно-временной точностью.

Более наглядно последнее заключение проиллюстрировано в [Mallat, 2002] и [Addison, 2002] в контексте принципа неопределенности Гейзенберга, согласно которому можно сказать, что чем сильнее Фурье или вейвлет базис локализован во времени, тем шире его спектр и наоборот. Поэтому результат любого спектрального преобразования имеет ограниченное частотно-временное разрешение и спектральные свойства современной музыки не укладываются в эти ограничения. Например, для разделения в спектре двух частот 27.5 Гц и 29.1 Гц, соответствующих двум наименьшим по высоте нотам фортепиано A0 и A#0, потребуется Фурье преобразование с размером окна 0.6 секунд. При применении в этом случае комплексного вейвлета Морле с областью определения, равном 8-ми, анализируемым частотам будут соответствовать масштабы вейвлета с временными длительностями 0.29 и 0.27 секунд и частотными интервалами 3.4 Гц и 3.7 Гц. Таким образом, элементы базиса Фурье, отличающиеся слабой локализацией, приводят к нежелательному размазыванию временного спектра, а компактность носителя вейвлет базиса является причиной недостаточного частотного разрешения для выделения фундаментальных частот музыкального строя. Отсюда можно сделать вывод, что оптимальной спектральной картины можно достичь только подстройкой размера окна в Фурье преобразовании или степени компактности вейвлет базиса в вейвлет преобразовании под спектральные свойства музыки.

Поэтому в данной работе была поставлена цель, разработать методологию для поиска компромиссного частотно-временного разрешения в алгоритмах спектрального анализа музыки. Для достижения этой цели и для синтеза эталонных музыкальных спектров будет использован подход на основе спектрального моделирования звука, который является достаточно эффективным методом для моделирования и синтеза музыки [Bonada et al., 2011]. В качестве эксперимента согласно разработанной методологии будет проведено исследование оптимального размера окна в оконном преобразовании Фурье и оптимальной компактности материнского вейвлета в вейвлет преобразовании для базы данных с классической музыкой.

Предлагаемый метод

Предлагаемый метод включает в себя три этапа (Рис. 1). Первый – спектральное моделирование музыкальных звуков, главная задача которого – представить необходимые нотные звуки музыкальных инструментов в виде спектральных характеристик, которые затем используются на втором этапе для синтеза музыкального звука и спектра, согласно заданному музыкальному тексту. Далее на последнем этапе производится поиск оптимальных параметров для исследуемого алгоритма спектрального анализа музыки путем минимизации расстояния между огибающими функциями синтезированного спектра и огибающими функциями спектра, полученного в результате спектрального анализа.

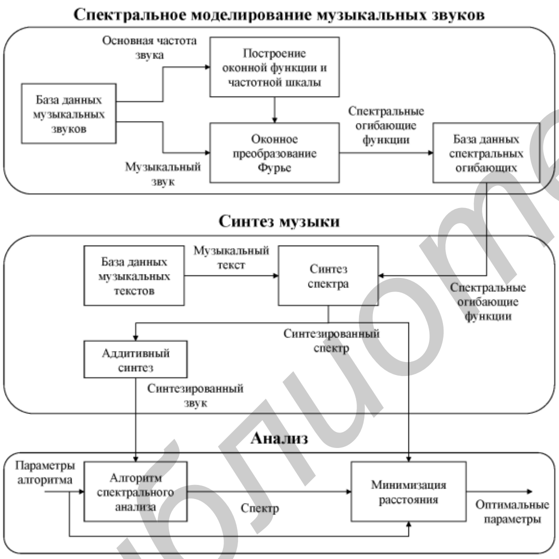

Рисунок 1 – Общая схема метода для поиска компромиссного частотно временного разрешения в алгоритмах спектрального анализа музыки

Здесь в первую очередь стоит отметить, что в базе данных музыкальных звуков должны содержаться все записи нот музыкальных инструментов, описанные в музыкальных текстах. Кроме этого для возможности синтеза и сравнения спектров необходимо, чтобы на каждом из трех этапов метода использовалась одна и та же частотная шкала.

Для моделирования принципов звукообразования простейших музыкальных звуков был использован метод спектрального моделирования музыки, суть которого заключается в извлечении из естественных звуков характерных для музыки спектральных огибающих функций с помощью оконного преобразования Фурье:

$$F(\tau, f) = \int_{\tau-T/2}^{\tau+T/2} s(t) w\left(\frac{t-\tau}{T}\right) e^{-2i\pi f(t-\tau)} dt, \qquad (1)$$

где $F(\tau, f)$ - функция спектральной огибающей; τ и f - момент времени и частота преобразования соответственно; $s(t)$ – анализируемый сигнал; $w(t)$ – оконная функция; T – ширина оконной функции.

Оконная функция и множество частот преобразования строятся исходя из значения основной частоты анализируемого звука. Согласно музыкальному равномерно темперированному строю возможные значения основной частоты удовлетворяют следующей формуле:

$$f'(n) = 440 \cdot 2^{(n-57)/12}, \qquad (2)$$

где $n \geq 0$ – порядковый номер ноты в музыкальном строе.

Значения гармонических частот музыкального звука основываются на целочисленных соотношениях с основной частотой и могут быть вычислены следующим образом:

$$f''(k) = kf_0, \qquad (3)$$

где $k \geq 1$ – порядковый номер гармоники; f_0 – основная частота звука.

Однако только первые двадцать гармоник в спектре звука являются наиболее значительными [Алдошина и др., 2006]. Отсюда можно заметить, что разница между любой из первых двадцати гармонических частот и ближайшей к ней частоте из музыкального строя не существенна. Если этим пренебречь, тогда, используя только частоты музыкального строя, допустимо анализировать и музыкальную тональность, и гармоники звука. Таким образом, частоты преобразования (и частотная шкала всего метода) могут быть заданы следующим образом:

$$\begin{cases} f(f_0, k) = 440 \cdot 2^{(n-57)/12} \\ n = \arg\min\limits_{n'} \left| kf_0 - 440 \cdot 2^{(n'-57)/12} \right| \\ 1 \leq k \leq 20 \end{cases} \qquad (4)$$

В качестве оконной функции была выбрана оконная функция Таки:

$$w(x) = \begin{cases} \sin^2(2\pi x), & 1/4 < |x| \leq 1/2 \\ 1, & |x| \leq 1/4 \end{cases} \qquad (5)$$

Благодаря плоской вершине она слабо локализует сигнал во временной шкале, что позволяет получить высокую четкость спектра в частотной шкале. Кроме этого в отличие от остальных оконных функций высокого разрешения (например, прямоугольного окна) окно Таки имеет незначительное влияние боковых лепестков для извлечения гармонических частот музыкальных звуков.

Для вычисления оптимальной ширины оконной функции оконное преобразование Фурье следует рассмотреть в качестве полосопропускающего фильтра. Тогда ширина окна Таки задается следующим образом:

$$T = \frac{1}{0{,}75 \cdot \Delta f_{0,5}}, \qquad (6)$$

где $\Delta f_{0,5}$ – ширина частотной полосы по амплитуде, равной 0,5.

Т.к. расстояние между гармоническими частотами в спектре музыкального звука равно основной частоте, поэтому для достижения максимально возможной временной четкости спектральных огибающих функций и для минимизации возможных отклонений гармонических частот спектра от ожидаемых частот можно принять $\Delta f_{0,5} = f_0$.

На этапе синтеза осуществляется синтез спектра и звука музыкальных композиций. Ключевыми источниками информации для осуществления этого процесса является база данных музыкальных текстов, содержащая музыкально-лингвистические свойства музыкальных композиций, и база данных спектральных огибающих, отражающая спектральные свойства каждой ноты музыкального текста. Аналитически синтез спектра может быть представлен следующим образом:

$$Spec(t, f) = \sum_{n \in N_t} (F_n(t - \tau_n, f) \cdot E_n(t - \tau_n)), \qquad (7)$$

где $Spec(t, f)$ – функция синтезированного спектра; N_t – множество нот, проигрываемое в момент времени t; F_n – огибающая спектра для ноты n; τ_n – момент времени в спектре, когда нота n начинает воспроизводиться; $E_n(t)$ – огибающая функция, задающая громкость воспроизведения и затухание ноты n:

$$E_n(t) = \begin{cases} a, & 0 \le t \le d \\ a - (t-d)/d', & d < t \le d + d' \end{cases} \qquad (8)$$

где $0 \le a \le 1$ – громкость воспроизведения; d и d' – длительности воспроизведения и затухания ноты.

Далее с помощью аддитивного синтеза воссоздается звуковой сигнал музыкального произведения:

$$S(t) = \sum_f (Spec(t, f) \cdot \sin(2\pi f t)). \qquad (9)$$

В процессе анализа производится минимизация расстояния между синтезированным спектром и спектром, извлеченным из синтезированного звука согласно следующей формуле:

$$p = \arg\min_{p'} D(Spec, T(S, p')), \qquad (10)$$

где p – оптимальное значение исследуемого параметра преобразования; D – функция расстояния; T – преобразование, которое лежит в основе исследуемого алгоритма спектрального анализа; $Spec$ и S – синтезированный спектр и сигнал музыкальной композиции соответственно.

В качестве функции расстояния D предлагается использовать манхэттенскую метрику. Она не содержит возведений в степень, что снижает влияние больших разностей на точность вычислений с плавающей запятой:

$$D = \sum_{t,f} \left| Spec'(t, f) - Spec''(t, f) \right|. \qquad (11)$$

Эксперимент

В ходе эксперимента на основе выше описанной методологии были исследованы пармаметры двух преобразований. Первый из них – размер окна в оконном преобразовании Фурье с окном Гаусса:

$$w(x) = e^{-x^2/0{,}045} \qquad (12)$$

И второй – компактность вейвлета в непрерывном вейвлет преобразовании на основе комплексного вейвлета Морле:

$$\psi(x) = e^{-(x l_0 / l)^2 / 2} * e^{-i 2\pi x}, \qquad (13)$$

где $l_0 = 10$ – ширина области определения вейвлета; l – компактность вейвлета (ширина временной локализации).

Таблица 1 – Список промоделированных акустических музыкальных звуков.

Инструмент	Звуки
Саксофон	D#3-F6
Банджо	C4-E7
Фагот	A#1-F5
Виолончель	C2-C6
Кларнет	D3-H6
Контрабас	C1-G4
Флейта	C4-D7
Гитара	E2-D5
Гобой	H3-G#6
Орган	C1-C6
Рояль	A0-C8
Тромбон	E2-E5
Труба	F3-D6
Туба	F1-F4
Альт	C3-C7
Скрипка	G3-G7

В качестве базы данных музыкальных текстов было использовано 610 30-секундных MID отрывков классической музыки. Также были промоделированы необходимые для синтеза звуки акустических музыкальных инструментов (Табл. 1). Все звуковые записи эксперимента хранились в формате WAV IEEE Float с частотой дискретизации 44100 Гц без сжатия.

Затем, используя каждое из двух исследуемых преобразований, из каждой синтезированной мелодии при заданном параметре преобразования извлекался спектр с временным шагом 10 мс в частотной шкале, согласно формуле (4). Таким образом, каждому музыкальному отрывку соответствовало оптимальное значение параметра преобразования. В результате для оконного преобразования Фурье было вычислено распределение оптимального размера окна в диапазоне от 50 мс до 1000 мс с шагом 50 мс, а для вейвлет преобразования – распределение оптимальной компактности вейвлета в диапазоне от 10 до 150 с шагом 5 (Рис. 2).

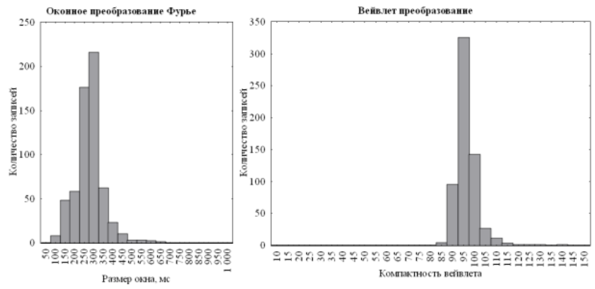

Рисунок 2 – Распределение значений оптимальных параметров преобразований.

На Рис. 2 заметна ярко выраженная концентрация значений размера окна вблизи значения 300 мс и значений компактности вейвлета вблизи значения 95. Следовательно, можно предположить, что при этих значениях в большинстве случаев исследуемые алгоритмы будут разделять фундаментальные частоты классической музыки с оптимальной частотно-временной точностью.

Заключение

Предложенный метод позволяет достигать оптимальной частотно-временной точности в алгоритмах спектрального анализа музыки. Этому способствует подход на основе спектрального моделирования музыкальных звуков. Результаты проведенного эксперимента с классической музыкой соответствуют гипотезе исследования об оптимальных значениях параметров преобразований. Следовательно, целесообразно дальнейшее развитие предложенного метода, путем моделирования музыки других жанров, а также путем исследования других типов преобразований.

Благодарности

Автор выражает особую благодарность музыкальной школе Аризонского университета, а так же организаторам проектов «Sound Exchange», «Composition Today» и «Midiworld» за безвозмездно предоставленные музыкальные тексты и звуковые записи акустических инструментов. Автор благодарит своего научного руководителя Хейдорова И.Э. за консультацию в проведении эксперимента.

Библиографический список

[Gómez et al., 2003] E. Gómez. Melody description and extraction in the context of music content processing / E. Gómez, A. Klapuri, and B. Meudic // Journal of New Music Research. – 2003 – V. 32(1). P. 23-40.

[Peeters, 2004] G. Peeters. A large set of audio features for sound description (similarity and classification) in the CUIDADO project / G.Peeters // Technical report, UIDADO IST Project. – 2004.

[Kronald-Martinet et al., 1987] R. Kronald-Martinet. Analysis of sound patterns through wavelet transform / R. Kronald-Martinet, J. Morlet, A. Grossman // International Journal of Pattern Recognition and Artificial Intelligence. – 1986. – V. 1. – № 2. – P. 273-302.

[Brown, 1991] Brown J.C. Calculation of a constant q spectral transform / Brown J.C. // Journal of the Acoustical Society of America. – 1991. – V. 89. – № 1. – P. 425-434.

[Mallat, 2002] Stéphane G. Mallat. A Wavelet Tour of Signal Processing / Stéphane G. Mallat // Academic Press. – 2009. – P. 23-34.

[Addison, 2002] Paul S. Addison. The Illustrated Wavelet Transform Handbook / Paul S. Addison // Taylor & Francis Group. – 2002. – P. 35-51.

[Bonada et al., 2011] Bonada J. Spectral Processing / Bonada J., Serra X., Amatriain X., Loscos A. // Book Chapter in DAFX Digital Audio Effects. – 2011. – P. 393-446.

[Алдошина и др., 2006] И. Алдошина. Музыкальная акустика / И. Алдошина, Р. Приттс // Учебник. СПб.: Композитор. – 2006.

OPTIMIZATION OF SHORT-TIME FOURIER AND CONTINUOUS WAVELET TRANSFORMS FOR SPECTRAL ANALYSIS OF MUSICAL SIGNALS

R.M. Aliyev

Belarusian State University, Minsk, Republic of Belarus

RomanAliyev@gmail.com

The method to improve time-frequency accuracy during spectral analysis of musical signals is offered. The method optimizes the parameters of spectral transforms and includes three stages – spectral modeling of sound samples, synthesis of music spectrums, and comparison of the synthesized and computed spectrums. Experiments with classical music databases presented the optimum window size in short-time Fourier transform and the optimum wavelet compactness in continuous wavelet transform are explored.

Key words: music recognition, spectral analysis, short-time Fourier transform, continuous wavelet transform.