

Санкт-Петербургский государственный университет

Кафедра теории управления

Землянова Анна Андреевна

Выпускная квалификационная работа магистра

**Распознавание полифонических музыкальных
звуковых потоков и получение нотной записи**

Направление 010302

Прикладная математика и информатика

Заведующий кафедрой,
доктор физ.-мат. наук,
профессор

Жабко Алексей Петрович

Научный руководитель,
кандидат физ.-мат. наук,
доцент

Ковшов Александр Михайлович

Рецензент

Козлов Максим Васильевич

Санкт-Петербург

2020

Содержание

1	Введение	3
2	Постановка задачи	5
3	Обзор литературы	6
4	Глава 1 Выделение основного тона	8
4.1	Метод «Сглаживания и усиления»	8
4.2	Мультисигнальная классификация	13
5	Метод на основе даунсэмплинга для определения смены частоты	16
6	Программная реализация	21
6.1	Общие положения	21
6.2	Описание интерфейса	21
7	Результаты	24
8	Выводы	26
9	Заключение	27
10	Список литературы	28

1 Введение

Природа человеческого мозга такова, что он постоянно находится в процессе обмена информацией с окружающей средой либо на этапе её анализа. Мозг вынужден непрерывно принимать информацию, приходящую с сенсорных рецепторов организма. Несмотря на то, что это, очевидно, энергозатратный процесс, изучение нового провоцирует выработку гормонов счастья, что позволяет распознать эту активность как эволюционно значимую.

Музыкальную информацию, в свою очередь, можно назвать одной из самых ценных в контексте производимого на нас влияния. Прослушивание музыки приводит к выработке дофамина – нейротрансмиттера, который играет не последнюю роль в нашем когнитивном, эмоциональном и поведенческом функционировании. Удивительным образом мозг распознаёт прослушивание структурированной последовательности звуков как приятный и полезный опыт.

Однако музыкальная информация в силу своей сложной структуры достаточно трудоёмка с точки зрения процессов её сохранения и обмена. В век информационных технологий налажены процессы оцифровки и передачи музыки в дискретном виде, поэтому её можно транслировать не только от человека к человеку в виде аудио и нотной информации, но и между электронно вычислительными устройствами и от них человеку. Тем не менее для возможности воспроизведения мелодии или аккордовой последовательности музыкантом необходимо представление музыки в виде, удобном для чтения.

Нотная запись – это специфический способ записи звукового материала с помощью письменных знаков. В отличие от электронных средств транслирования

музыкальной информации нотация передаёт специфику музыкальной логики, что делает её незаменимой с точки зрения постижения человеческим сознанием. Самые распространенные виды нотации: буквенная, линейная и табулатура.

2 Постановка задачи

Целью представленной работы является разработка алгоритма и создание программного продукта, предназначенных для получения нотного представления музыкального оцифрованного аудиосигнала, в том числе сигнала, содержащего полифонию.

Подзадачи:

- Выделение одного или нескольких основных тонов в аудиосигнале в каждый момент времени.
- Определения моментов смены частот основного тона
- Программная реализация

3 Обзор литературы

В пособии Ричарда Лайонса [6] читателю предлагается подробное и широкоохватное описание методов цифровой обработки сигналов. Здесь рассмотрены как самые базовые понятия в данной сфере (дискретное и быстрое преобразование Фурье) так и более узконаправленные (цифровая фильтрация, преобразование частоты дискретизации). Книга не только ссылается на необходимый для понимания методов математический аппарат, но и содержит всё требующееся математическое обоснование (статистика, преобразование Лапласа и др.).

Ресурс [2] в основном посвящен задаче детекции основного тона в музыкальных и речевых сигналах. Как известно, подходов в данной проблеме существует немало (выделение во временной области, выделение в спектральной области, алгоритмы на основе вейвлетов, кепстральный метод), и в этом пособии описано подавляющее большинство. Помимо этого, здесь можно ознакомиться с наиболее популярными проблемами, возникающими при работе с определением несущей частоты (такие как "ошибка октав") и их возможными решениями.

Работа [9] освещает DOA Estimation методы (методы определения направления прихода сигналов)

Вариация применения алгоритма мультисигнальной классификации рассмотрена в статье [8]. Здесь вместо спектрального разложения матрицы корреляции предлагается использование быстрого преобразования Фурье.

В книге [3] освещаются проблемы работы с аудиосигналами речевого типа, рассматриваются вопросы анализа голосовых сигналов с целью совершенство-

вания процессов управления электронно вычислительными машинами с помощью речи. Изучаются как общие для звуковых сигналов вопросы (создание цифровых фильтров, квантование) так и специфические аспекты анализа речи.

4 Глава 1 Выделение основного тона

В этом разделе рассмотрим задачу получения высоты частоты основного тона рассматриваемого сигнала. Этот вопрос играет первостепенную роль в процессе получения представления звукового потока в виде нотного представления. Одним из самых распространённых методов решения проблемы поиска частоты в любом периодическом сигнале, без сомнения, является Фурье анализ, однако при работе с аудиосигналом и тем более при необходимости получения частоты с точностью достаточной для отнесения частоты к определенной ноте, Фурье-анализ не всегда работает достаточно хорошо из-за проблемы спектральной утечки, а также невозможности оценить частоту с точностью выше, чем один сэмпл.

4.1 Метод «Сглаживания и усиления»

Представленный метод позволяет в каждый момент времени получать значение значение одной или нескольких основных частот присутствующих в аудиосигнале. Он основывается на гипотезе о том, что любой периодический сигнал стремится к гармоническому, при применении к исходному сигналу достаточного количества операций сглаживания функцией Стеклова и умножения на коэффициент (чтобы не допустить полного исчезновения сигнала)

Рассмотрим простейшую синусоидальную модель сигнала: $r \sin(kt)$

r – амплитуда; k – частота; t – время.

Применяя к нему операцию сглаживания функцией Стеклова окном шириной $2a$ получим:

$$\begin{aligned}
\frac{1}{2a} \int_{t-a}^{t+a} r \sin(k\tau) d\tau &= \frac{r}{2ak} \int_{t-a}^{t+a} \sin(k\tau) dk\tau = -\frac{r}{2ak} (\cos k(t+a) - \cos k(t-a)) = \\
&= -\frac{r}{2ak} \left(\cos(kt) \cos(ka) - \sin(kt) \sin(ka) - \right. \\
&\quad \left. - \cos(kt) \cos(ka) + \sin(kt) \sin(ka) \right) = \\
&= \frac{r}{2ak} 2 \sin(kt) \sin(ka) = \frac{r \sin(ka)}{ak} \sin(kt) = \frac{\sin(ka)}{ka} r \sin(kt) = G \cdot r \sin(kt),
\end{aligned}$$

где «коэффициент сглаживания» $G = \frac{\sin(ka)}{ka} \leq 1$.

Для сигнала состоящего из нескольких чистых гармоник, имеющих разную частоту, последовательное сглаживание оставит одну гармонику наиболее низкой частоты.

Действительно: $A(t) = p \sin ut + q \sin vt$, где p и q — амплитуды, а u и v — частотные коэффициенты, пусть $u < v$. Сглаживая по Стеклову, имеем:

$$\begin{aligned}
\frac{1}{2a} \int_{x-a}^{x+a} (p \sin u\tau + q \sin v\tau) d\tau &= \frac{1}{2a} \left(p \int_{x-a}^{x+a} \sin u\tau d\tau + q \int_{x-a}^{x+a} \sin v\tau d\tau \right) = \\
&= \frac{\sin ua}{ua} p \sin ut + \frac{\sin va}{va} q \sin vt = G(a, u) p \sin ut + G(a, v) q \sin vt.
\end{aligned}$$

Отсюда, после сглаживания получается сигнал, состоящий из суммы тех же гармоник, но домноженных на коэффициенты сглаживания $G(a, u)$ и $G(a, v)$ соответственно. Если a такое, что $G(a, u) > h$, то $G(a, u) > G(a, v)$. Последовательно сглаживая n раз и, домножая после каждого сглаживания сигнал на $G(a, u)^{-1}$, получим из сигнала $A(t)$ сглаженный сигнал $A_n(t)$ порядка n

$$A_n(t) = p \sin ut + \left(\frac{G(a, v)}{G(a, u)} \right)^n q \sin vt.$$

Так как $G(a, u) > G(a, v)$, то

$$\left(\frac{G(a, v)}{G(a, u)}\right)^n \xrightarrow{n \rightarrow \infty} 0, \quad \text{и} \quad \lim_{n \rightarrow \infty} A_n(t) = p \sin ut.$$

Таким образом, после последовательных усреднений с усилением в пределе останется синусоида наибольшего периода. При этом исходный сигнал может состоять из любого числа гармоник.

Отсюда следует, что любой периодический сигнал стремится к гармоническому при стремлении числа усреднений и умножений на коэффициент (для того, чтоб сигнал не угас при сглаживании) к бесконечности.

Таким образом, алгоритм метода будет выглядеть так:

1. Вычисляется среднеквадратичная амплитуда сигнала, при передвижении по нему полосой в 1000 отсчётов (значение выбирается из расчёта минимальной длительности одной частоты).

$$P_t = \sqrt{\frac{1}{N} \sum_{-N/2}^{i=N/2} x_{t-i}^2} \quad t = \overline{0, (T-1)}$$

t — номера отсчётов, T — число отсчётов, T/S — длительность рассматриваемой звукозаписи в секундах, x_t — отсчёты сигнала, N — количество отсчетов сигнала в окне.

2. Сигнал усредняется полосой 20 отсчётов ($a = 10$ отсчётов).

$$x_t^{[1]} = \frac{1}{2a} \sum_{i=t-a}^{t+a} x_i$$

,

3. Снова вычисляется среднеквадратичная амплитуда полосой 1000 отсчётов для усредненного массива сэмплов сигнала

$$P_t^{[1]} = \sqrt{\frac{1}{N} \sum_{-N/2}^{i=N/2} (x_{t-i}^{[1]})^2} \quad t = \overline{0, (T-1)}$$

где $[1]$ – кратность сглаживания.

4. Вычисляются коэффициенты сглаживания (ослабления) сигнала G_t , как отношение усреднённой и исходной усреднённых амплитуд в каждом отсчёте .

$$G_t = P_t^{[n]} / P_t^{[0]}$$

5. . Сигнал умножается на $1/G_t$ в каждом отсчёте, чтобы восстановить амплитуду.

$$x_i * \frac{1}{G_t}$$

6. Последовательность действий 1 – 5 повторяется до тех пор, пока G_t не устоится, то есть, если два последовательных сглаживания с усилением дают для всех отсчётов одни и те же G_t то это означает, что остались чистые синусоиды низших частот.

7. После этого запускается алгоритм поиска отрезков с постоянной частотой, а именно находятся отрезки с постоянными G_t , а также отрезки, где G_t плавно (линейно) переходит из одного в другое. Длина таких переходных отрезков равна удвоенной ширине полосы усреднения. В нашем случае —

2000. Середина переходного отрезка принимается за границу соседних длительностей постоянной частоты.

8. Каждая длительность постоянной частоты дальше будет обрабатываться отдельно.
9. Последовательным поиском подбирается такая полоса усреднения $2a$, для которой $G_i = 2/\pi$. В этом случае программа выдаёт точное значение частоты, которое и так можно вычислить, поскольку период равен $4a$, следовательно, частота равна $0,25/a$.

При анализе сигнала, содержащего более чем один основной тон в некоторый момент времени (полифонию), найденная гармоника отнимается от исходного сигнала, давая возможность продолжить поиск более высоких частот

4.2 Мультисигнальная классификация

Данный алгоритм оценивает частотную составляющую аудиосигнала, используя собственные вектора автокорреляционной матрицы сигнала.

Метод работает в предположении, что сигнал $x(n)$ состоит из p комплексных экспонент с добавлением белого гауссовского шума.

Рассмотрим сигнал:

$$x(n) = \sum_{i=1}^M A_i \cos(2\pi f_i n + \phi_i) + e(n)$$

где:

A_i – амплитуда; f_i – частота ϕ_i – фаза

$e(n)$ – шум с нулевым средним (белый шум)

M – номер гармоники

Его можно представить в следующем виде :

$$x(n) = \sum_{i=1}^M A_i e^{j(2\pi f_i n + \phi_i)}$$

Ковариационная функция от сигнала $x(n)$ тогда будет выглядеть:

$$r(k) = \sum_{i=1}^M a_i e^{j(2\pi f_i k)} + P_n \delta(k)$$

$a_i = A_i^2$, P_n – дисперсия $e(n)$

Рассмотрим R – матрица автокорреляции размерности $P \times P$.

$$R = \begin{bmatrix} r(0) & r(1) & \dots & r(P-1) \\ r(-1) & r(0) & \dots & r(P-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(P-1) & r(P-2) & \dots & r(0) \end{bmatrix}$$

Матрица R – матрица Теплица. В соответствии с её свойствами

$$\text{rank}(R) = M$$

В соответствии со спектральной декомпозицией автокорреляционной матрицы R , можно записать:

$$R = \sum_{k=1}^P \lambda_k e_k e_k^H$$

где λ_k и e_k -собственные числа и соответствующие собственные вектора.

$(\cdot)^H$ – оператор Эрмитава транспонирования.

Допустим, будем сортировать собственные числа в порядке убывания, тогда матрицы

$$V_s = [V_1, V_2, \dots, V_M] \quad V_N = [V_{M+1}, V_{M+2}, \dots, V_P]$$

содержат в себе собственные вектора отвечающие подпространству сигнала и подпространству шума соответственно.

На практике матрицу автокорреляции находят как :

$$\hat{R} = \frac{1}{K} \sum_{k=1}^P x(n)x(n)^H$$

Её спектральная декомпозиция:

$$\hat{R} = \hat{V}_s \hat{\Lambda}_s = \hat{V}_s^H = + \hat{V}_N \hat{\Lambda}_N \hat{V}_N^H$$

Где собственные векторы опять же сортированы в порядке убывания ($\lambda_1 \geq \lambda_2 \dots \geq \lambda_P$) И в матрицах

$$\hat{V}_s = [\hat{V}_1, \hat{V}_2, \dots, \hat{V}_M] \quad \hat{V}_N = [\hat{V}_{M+1}, \hat{V}_{M+2}, \dots, \hat{V}_P]$$

содержатся столбцы собственных векторов подпространства сигнала и подпространства шума соответственно.

Диагональные матрицы

$$\hat{\Lambda}_S = \text{diag}[\lambda_1, \dots, \lambda_M] \quad \hat{\Lambda}_N = \text{diag}[\lambda_{M+1}, \dots, \lambda_P]$$

содержат на главной диагонали собственные числа из подпространства сигнала и подпространства помехи соответственно.

Собственные вектора V_i , соответствующие первым p сортированным в порядке убывания собственным числам λ_i матрицы R_x охватывают подпространство сигнала.

Оставшиеся $n - p$ собственных векторов соответствуют ортогональному пространству помехи.

Рассмотрим опорный вектор

$$a(\omega) = [1 \ e^{-j\omega} \ \dots \ e^{-j(p-1)\omega}] \quad \omega = 2\pi f_i$$

Учитывая его ортогональность подпространство помехи, получаем:

$$a^H(\omega)V_nV_n^H a(\omega) = 0$$

Для данного алгоритма функция оценки частоты:

$$P(e^{j\omega}) = \frac{1}{\sum_{i=p+1}^n |e^H v_i|^2},$$

где v_i – собственные векторы шума

$$e = [1 \ e^{j2\omega} \ \dots \ e^{j(n-1)\omega}]$$

Положение p самых высоких пиков функции оценки частоты будут давать оценки частот, присутствующих в сигнале.

5 Метод на основе даунсэмплинга для определения смены частоты

Будем рассматривать синусоидальную модель аудио сигнала

$$x[n] = \sum_{r=1}^R A_r[n] \cos(2\pi f_r n),$$

где

$x[n]$ – дискретные отсчёты сигнала

R – общее число гармонических синусоид входящих в состав рассматриваемого сигнала и изменяющихся во времени.

n – временная составляющая

$A_r[n]$ – мгновенная амплитуда

f_r – мгновенная частота

Применение к сигналу операции намеренного снижения качества или, иначе, даунсэмплинга, позволяет получить приблизительную оценку модели дискретного сигнала, теоретически получившегося бы при использовании более низкой частоты дискретизации.

$$x[Mn] = x[n] \downarrow M$$

где

$x[n]$ – входные отсчеты сигнала до децимации

M – длительность строба.

Далее будем рассматривать двумерную структуру размерности $(n \times 2m + 1)$

$$f(x, y) = \begin{bmatrix} f(m; 0) & f(m; 1) & f(m; 2) & \dots & f(m; n-1) \\ f(m-1; 0) & f(m-1; 1) & f(m-1; 2) & \dots & f(m-1; n-1) \\ \vdots & \vdots & \vdots & & \vdots \\ f(1; 0) & f(1; 1) & f(1; 2) & \dots & f(1; n-1) \\ f(0; 0) & f(0; 1) & f(0; 2) & \dots & f(0; n-1) \\ f(-1; 0) & f(-1; 1) & f(-1; 2) & \dots & f(-1; n-1) \\ \vdots & \vdots & \vdots & & \vdots \\ f(-m+1; 0) & f(-m+1; 1) & f(-m+1; 2) & \dots & f(-m+1; n-1) \\ f(-m; 0) & f(-m; 1) & f(-m; 2) & \dots & f(-m; n-1) \end{bmatrix}$$

n – количество сэмплов сигнала после применения операции децимации

m – максимальное значение амплитуды сигнала в уменьшенном масштабе
как графическое представление полученного ранее дискретного аудиосигнала.

Вектор значений амплитуд исходного дискретного разреженного сигнала будем обозначать:

$$A = (a_1, a_2, \dots, a_n).$$

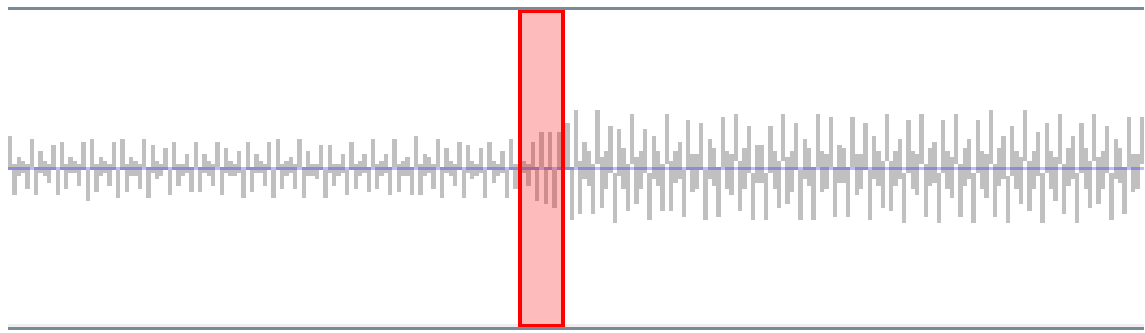
n – количество сэмплов после прореживания.

Заполнение матрицы будет осуществляться следующим образом:

Для $\forall j$, если $a_j \geq 0$ и $0 \leq i \leq a_j$ и либо $a_{j-1} < i$, либо $a_{j+1} < i$, $\rightarrow f(i, j) = 1$

Для $\forall j$, если $a_j \leq 0$ и $0 \geq i \geq a_j$ и либо $a_{j-1} > i$, либо $a_{j+1} > i$, $\rightarrow f(i, j) = 1$

Все остальные ячейки заполняются нулями



Графическое изображение децимированного сигнала. В красном прямоугольнике – момент сметы частоты

Теперь, пробегая по двумерному массиву окном высотой n и шириной k (ширина окна выбирается в зависимости от специфики данных), для каждого окна подсчитываем число постоянства значений f для каждой строки.

При этом составляется матрица G размерности $2m + 1 \times s$,

$$G = \begin{bmatrix} g(0; 0) \dots g(0; s - 1) \\ g(1; 0) \dots g(1, s - 1) \\ \vdots \\ g(2m; 0) \dots g(m; k) \end{bmatrix}_{2m+1 \times s}$$

где $s \in N$, s – максимальная длина интересующего нас набора непрерывных последовательностей единиц в матрице f .

Так, то для $\forall g(i, j)$ равно числу непрерывных последовательностей единиц в матрице f на i -й строке.

Пример:

Пусть $k = 50$ – длина полосы. $i = 7$, $s = 5$, рассматриваем отрезок седьмой строки матрицы f , попадающий в окно ширины k –

Например: $[00011\ 01110\ 00011\ 01110\ 00011\ 01110\ 00011\ 01110\ 00011\ 01110]$, тогда:

$$g(7, 1) = 0$$

$$g(7, 2) = 5 \text{ – Пять раз встречается последовательность "11"}$$

$$g(7, 3) = 5 \text{ – Пять раз встречается последовательность "111"}$$

$$g(7, 4) = 0$$

$$g(7, 5) = 0$$

Далее для каждой из полученных l матриц $G[i], i \in [1 \dots l]$

где

$$l = \frac{n}{k}$$

l - количество окон.

Выполняем операцию векторизации для каждой из матриц $G[i]$, например:

$$\begin{bmatrix} g(0; 0) \dots g(0; s - 1) \\ g(1; 0) \dots g(1, s - 1) \\ \vdots \\ g(2m; 0) \dots g(2m; s - 1) \end{bmatrix} \rightarrow$$

$$\rightarrow V_1 = (g(0; 0) \dots g(0; s - 1), g(1; 0) \dots g(1, s - 1), \dots, g(2m; 0) \dots g(2m; s - 1))$$

В результате имеем набор векторов

$$V_1, V_2, \dots, V_l,$$

где

$$l = \frac{n}{k}$$

l - количество окон.

Затем для каждого из векторов $V_1 = (v_{11}, \dots, v_{1k}), V_2 = (v_{21}, \dots, v_{2k}), \dots, V_l = (v_{l1}, \dots, v_{lk})$, вычисляем евклидову метрику:

$$m_i = \sqrt{v_{i1}^2 + v_{i2}^2 + \dots + v_{ij}^2}$$

где

$$i \in [1 \dots l] ; j \in [1 \dots k]$$

Таким образом, получаем ряд из l метрик, каждая из которых соответствует конкретному окну ширины k из исходной матрицы.

Для $\forall i \in [2 \dots l]$ находим $m_{i+1} - m_i$.

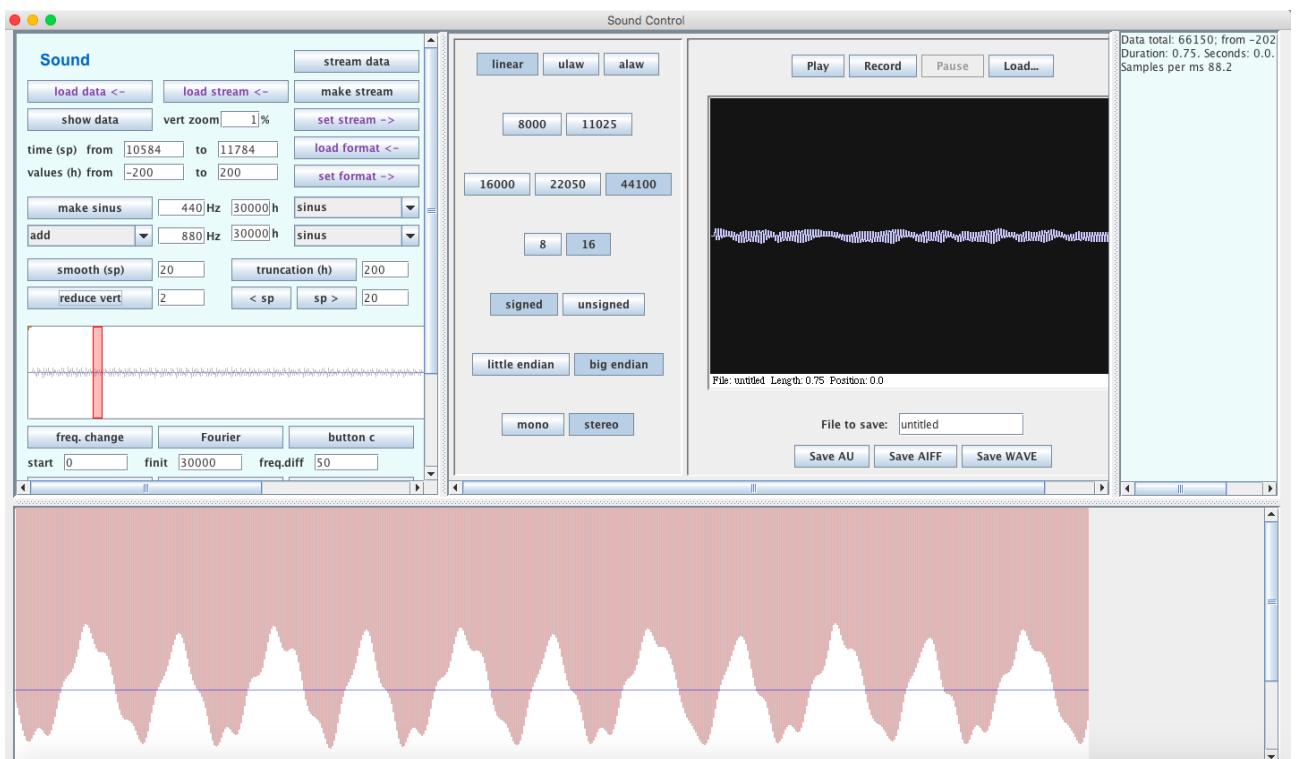
Из полученной одномерной структуры, состоящей из разностей метрик несложно получить информацию о номере отсчёта, соответствующего смене частоты в исходном дискретном аудиосигнале, отследив изменение в самой структуре.

6 Программная реализация

6.1 Общие положения

- Программный продукт реализован на языке программирования Java.
- Работа с графикой осуществлялась с помощью графической библиотеки Swing.
- Программа работает с аудиоданными несжатого формата WAV
- Используются классы и методы Java Sound API.

6.2 Описание интерфейса



В нижней части окна графически прорисовываются значения амплитуды анализируемого сигнала. В верхней части в правом окне осуществляется вывод

сообщений. В среднем окне верхней части встроены методы для воспроизведения и записи звука. Оно также разделено на два отсека. В левом отсеке можно группу кнопок, отвечающих за текущий формат представления звукового сигнала (моно/стерео; обратный порядок байтов при записи/ прямой порядок байтов при записи; запись со знаком/запись без знака; 8 бит на сэмпл/16 бит на сэмпл; частота дискретизации). В правом – кнопки проигрывания, записи, паузы и загрузки из файла, а также графическое поле и органы, управляющие форматом записи звука в файл.

В левой части верхнего окна располагаются кнопки, завязанные на методы для генерации и модификации текущего звукового сигнала. Тут основную роль играют три объекта, хранящих звуковые данные: `data` , `format` и `stream` .

`data` – массив сэмплов; `format` – формат записи звука, описанный в среднем окошке; `stream` – звуковой поток, то есть объект, позволяющий проиграть звук или записать его в файл. Кнопки с надписями фиолетового цвета предназначены для различных действий с объектом "звуковой поток (`stream`)".

"`make stream`" создает звуковой поток из текущего массива сэмплов в соответствии с установленным форматом записи звука.

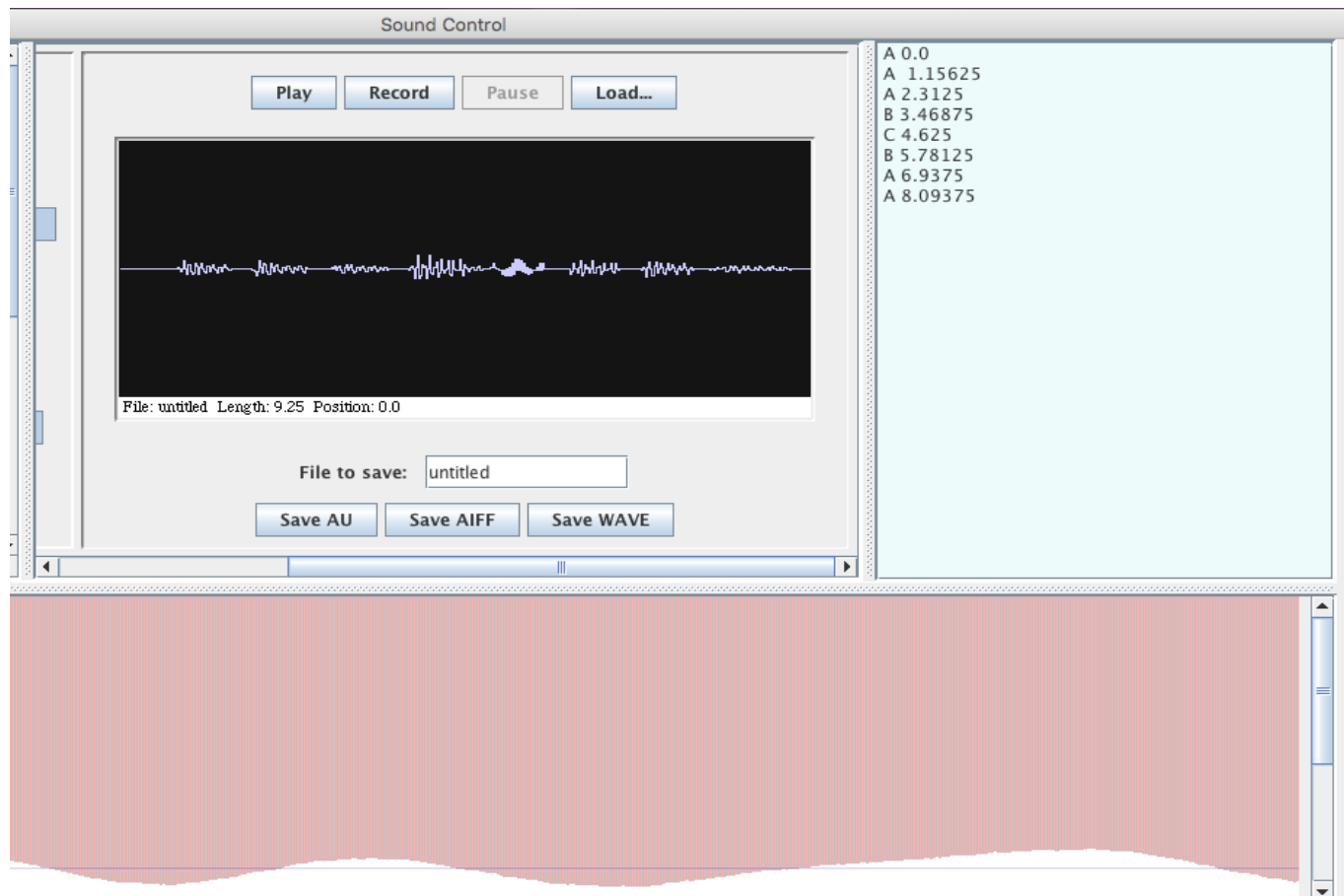
"`stream data`" получает отсчеты из текущего звукового потока (`stream`) в соответствии с текущим форматом записи звука.

"`show data`" графически отображает массив сэмплов в нижней части окна программы. Текстовые поля "`time from`" и "`to`" показывают с какого по какой номер элементов массива неделимых значений звука (`data`) отображаются в нижней части внешнего окна. Текстовые поля "`values from`" и "`to`" показывают границы по вертикали отображаемых значений звука в нижней части окна. Здесь также представлен метод «`make sinus`», генерирующий массив данных звука (синусоидальный, пилообразный, импульсная знакопеременная функция) Также под

этой кнопкой расположен метод с выбором параметров для наложения двух функций различными способами. Над графическим полем расположены три кнопки преобразования данных (сглаживание, обрезание значений) и две кнопки сдвига графического отображения сэмплов в нижней части внешнего окна. Кнопки под графическим полем выводят в диалоговое окно результат реализованных методов для анализа аудиосигнала. На "button" завязан метод мультисигнальной классификации. Кнопка "freq.change" вызывает метод, вычисляющий переходы через ноль данных сигнала. Кнопка "Fourier" выводит частоты, полученные с помощью применения преобразования Фурье к сегменту сигнала. Кнопка "av.sq.power" завязана на метод, который вычисляет среднеквадратичную амплитуду с бегущим окном шириной, указанной в соседнем поле "sp".

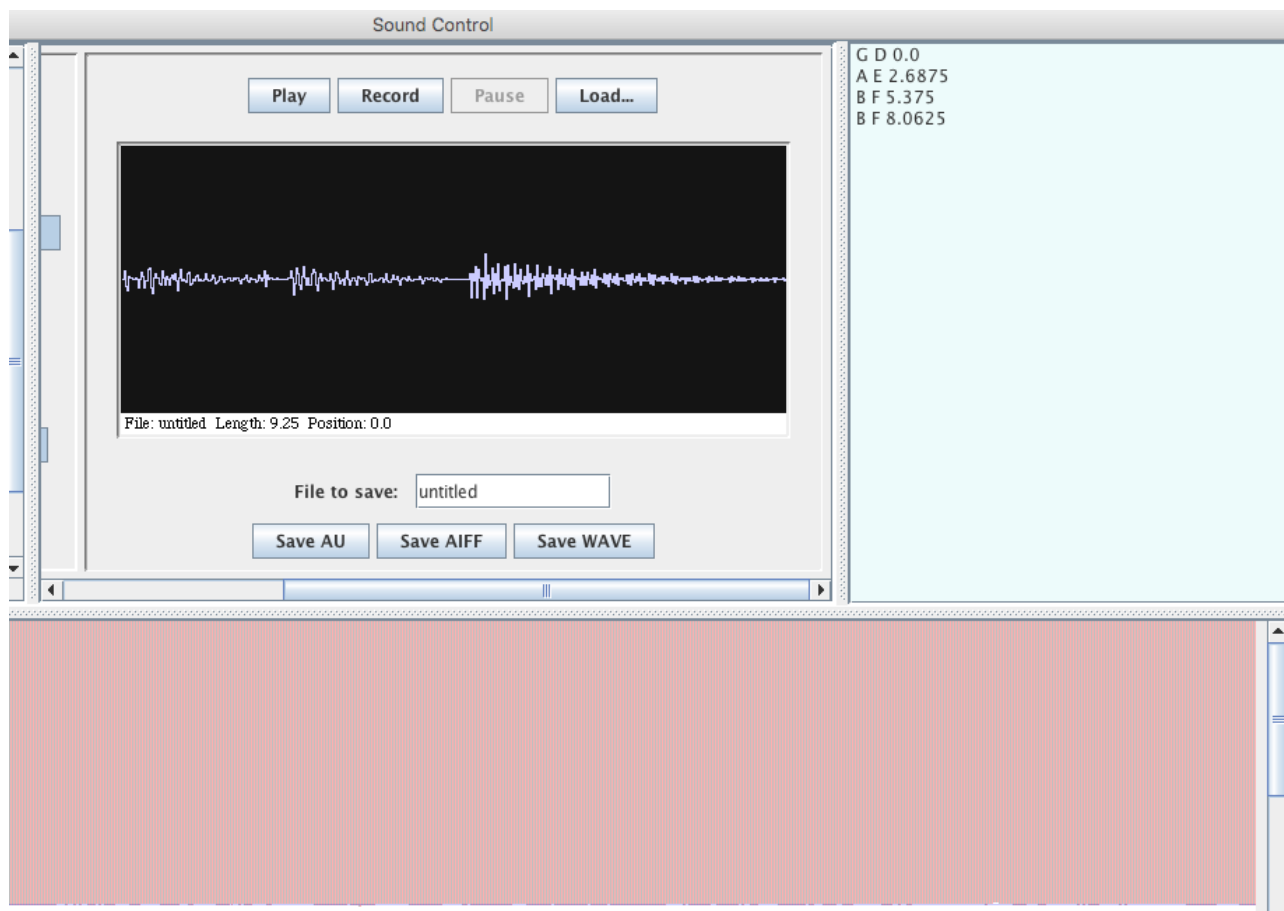
7 Результаты

1) Одноголосый сигнал длиной 815850 сэмплов, длительностью 9,25 секунд.



В окне для вывода текстовых сообщений: в первом столбце отображена буквенная нотация частот, присутствующих в сигнале, во втором – момент в секундах смены предыдущей частоты на текущую.

2) Полифонический сигнал длиной 948150 сэмплов, длительностью 10,75 секунд.



В окне для вывода текстовых сообщений: в первых двух столбцах отображены буквенные нотации частот, присутствующих в сигнале (в каждой строчке – 2 одновременно звучащие ноты), в третьем – момент в секундах смены предыдущих частот на текущие. Частоты, звучащие в сигнале в одни и те же моменты времени также синхронно сменяются. Сигнал представляет из себя 3 последовательно проигрываемых квинты.

8 Выводы

Мультисигнальная классификация работает лучше чем, к примеру, классические для задач поиска несущей частоты алгоритмы поиска пиков спектра в методах Дискретного преобразования Фурье или быстрого преобразования Фурье в условиях наличия шума и в случае, если заранее известно количество основных частот в определённый момент времени, потому что она использует информацию о количестве основных тонов в том числе для игнорирования шума. Однако Мультисигнальная классификация является крайне затратным по памяти алгоритмом даже при правильной сегментации сигнала.

Метод «сглаживания усиления» позволяет с высокой точностью находить частоты, присутствующее в сигнале в определённый момент времени при корректном подборе ширины сглаживающего окна, а также получать моменты смены частот. В коллаборации с методом на основе даунсэмплинга для определения смены частоты, он предоставляет возможность получить достаточно точную интерпретацию сигнала по соотношению частот по длительности, достаточно для получения нотного представления.

9 Заключение

Результатом представленной работы явилось создание программного продукта, предназначенного для анализа оцифрованного музыкального сигнала и выделения его нотации.

Для удобства проведения тестирования методов программа также снабжена алгоритмами генерации, модификации, записи и воспроизведения аудиосигнала.

На данный момент задача получения нотации из музыкального сигнала решается в основном коммерческими программными средствами с закрытым исходным кодом, такими как «Melodyne», «Transcribe».

10 Список литературы

[1] Оппенгейм А., Шафер Р., Цифровая обработка сигналов, М, «Техносфера» 2006.- 856с.

[2] Fast, Accurate Pitch Detection Tools for Music Analysis. Philip McLeod, University of Otago, Dunedin, New Zealand, 2008.

[3] Рабинер Л.Р., Шафер Р.В., Цифровая обработка речевых сигналов, М, «Радио и связь » 1981

[4] Noll, M. Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum, and a maximum likelihood estimate. In Proceedings of the Symposium on Computer Processing ing. Communications / M. Noll. – Polytechnic Institute of Brooklyn, 1969. – P. 779–797.

[5] Philip McLeod, Geoff Wyvill, A SMARTER WAY TO FIND PITCH, University of Otago, Department of Computer Science

[6] Ричард Лайонс. Цифровая обработка сигналов. 2006 год. 656 стр.

[7] Java Sound Programmer Guide

<https://docs.oracle.com/javase/8/docs/technotes/guides/sound/programmerguide/contents.html>

[8] FAST MUSIC – AN EFFICIENT IMPLEMENTATION OF THE MUSIC ALGORITHM FOR FREQUENCY ESTIMATION OF APPROXIMATELY

PERIODIC SIGNALS Orchisama Das, Jonathan S. Abel, Julius O. Smith III
Center for Computer Research in Music and Acoustics, Stanford University

[9] Kalmar Vaxjo. DOA estimation based on MUSIC algorithm

[10] The Open Electrical & Electronic Engineering Journal, 2015, 9, 38-42