



ный вопрос дидактического элемента, то ему предъявляется задание с более низким коэффициентом сложности и т.д. Граничные значения коэффициентов сложности описываются в используемой при диагностике модели.

Компьютерная интеллектуальная адаптивная система тестирования должна обладать следующим набором характеристик:

- Открытость и расширяемость.
- Нелинейность воспроизведения диагностического контента.
- Известная трудность.
- Универсальность диагностической модели.
- Достоверность и точность результатов адаптивного тестирования.

### Литература

1. Brusilovsky, P., Methods and techniques of adaptive hypermedia. User Modeling and User-Adapted Interaction, 6 (2-3), pp. 87-129
2. Brusilovsky, P., Adaptive and Intelligent Technologies for Web-based Education. In C. Rollinger and C. Peylo (eds.), Special Issue on Intelligent Systems and Teleteaching, Konstruktive Intelligenz, 4, pp. 19-25
3. Никифоров О.Ю. Использование адаптивных систем компьютерного тестирования // Гуманитарные научные исследования. 2014. № 4

А.О. Новиков

## РАСПОЗНАВАНИЕ НОТ В АУДИОФАЙЛЕ

(Самарский государственный аэрокосмический университет)

На данный момент существуют программы для распознавания названия музыки по ее записи, причем допускается, что на этой записи помимо самой музыки присутствуют посторонние звуки, да и качество самой записи может быть относительно невысоким. Например, *Shazam* – одна из таких программ. Она составляет отпечаток записи и по определенному алгоритму сравнивает с огромной внутренней базой имеющихся отпечатков на сходство, а затем выдает название песни. Это очень удобный инструмент. Но программ, которые бы распознавали именно ноты (возможно зажатых одновременно более одной), играемых инструментов, сейчас не существует (по крайней мере, в публичном доступе). Попытка разработки такой программы и является целью данного доклада, ограничившись, для начала, нотами фортепиано, что облегчает задачу отсутствием сложных артикуляций и определения типов инструментов.

*Аудиосигнал* – по сути, функция зависимости амплитуды от времени. В цифровом виде аудиосигнал дискретный, то есть в файле хранится последовательность значений амплитуды (отсчеты), взятые от непрерывного сигнала через равные промежутки времени. Частота дискретизации – количество таких отсчетов за одну секунду.

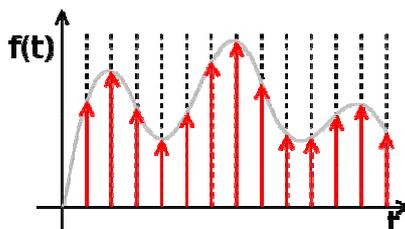


Рис. 1. Дискретизация сигнала

Такое представление сигнала позволяет увидеть основные события в звуке, такие как изменения в громкости или паузы, но обычно, если говорить об отдельных нотах, оно мало полезно, так как аудиосигнал обычно состоит из большой суммы разных синусоид. Поэтому по аудиосигналу составляется спектрограмма – зависимость амплитуды каждой составляющей сигнала от времени.

Составляться спектрограмма может по-разному. Например, используя преобразование Фурье. Это математический аппарат для разложения сигнала на синусоидальные колебания. Поскольку музыка не является статичным сигналом, ее спектр меняется во времени. Поэтому при спектральном анализе берется последовательность коротких фрагментов музыки, на которых сигнал может считаться практически статичным. Для анализа таких фрагментов цифрового аудиосигнала существует дискретное преобразование Фурье.

$$x(n) = \sum_{k=0}^{N/2} X_k \cos \frac{2\pi k(n + \varphi_k)}{N}$$

$X_k$  – амплитуда синусоидального колебания;

$\frac{kF}{N}$  – частота синусоиды, где  $F$  – частота дискретизации;

$N$  – число отсчетов исходного сигнала на анализируемом интервале.

Как видно из формулы, частоты синусоид, на которые раскладывается сигнал, равномерно распределены от 0 (постоянная составляющая) до  $F/2$  – максимально возможной частоты в цифровом сигнале. На практике применяют FFT (fast Fourier transform) – алгоритм быстрого вычисления дискретного преобразования Фурье. Стоит отметить, что FFT раскладывает сигнал не по тем частотам, которые на самом деле присутствуют в сигнале, а по фиксированной равномерной сетке частот. Например, если частота дискретизации равна 48 кГц и размер интервала FFT выбран 4096 отсчетов, то FFT раскладывает сигнал по 2049 частотам: 0 Гц, 11.72 Гц, 23.44 Гц, ..., 24000 Гц. Если частота тона совпадает с одной из частот сетки FFT, то спектр будет выглядеть "идеально": единственный острый пик укажет на частоту и амплитуду тона. Если же частота тона не совпадает ни с одной из частот сетки FFT, то он "соберёт" тон из имеющихся в сетке частот, скомбинированных с различными весами. График спектра при этом размывается по частоте. Такое размытие очевидно нежелательно.

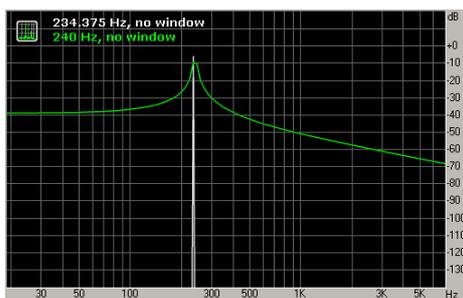


Рис. 2. Сравнение спектров при совпадении и несовпадении реальной частоты с сеткой частот FFT

Для уменьшения размытия спектра, сигнал перед вычислением FFT умножается на весовые окна – гладкие функции, спадающие к краям интервала. Несомненно, это одна из помех в распознавании нот. Другой из них являются гармоники – частотные составляющие, имеющие частоту в 2, 3, ... раза большую, чем основной тон. Они присутствуют в звучании ноты практически любого инструмента и амплитуда их не всегда ниже амплитуды основного тона.

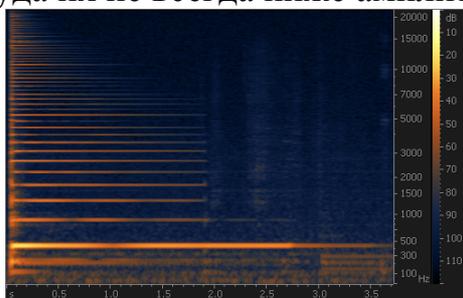


Рис. 3. Спектрограмма гитарной ноты А4 (440 Гц)

Еще одной проблемой распознавания является наложение гармоник нескольких нот, звучащих одновременно, друг на друга, что также затрудняет понять где находятся основные тона, а значит, определить играемые ноты. Для облегчения задачи при разработке алгоритма считалось, что основной тон всегда громче своих гармоник, что экспериментально оказалось верно, почти для всех нот фортепиано выше первой октавы.

Работа алгоритма заключается в следующем. По всему сигналу берется последовательность окон (спектров). Для данной мелодии устанавливается порог, который выбирается экспериментально. Частоты ниже этого порога удаляются, отсеивая, таким образом, лишние гармоники и некоторые шумы. Если на спектре есть составляющая, амплитуда которой выше этого порога, то эта составляющая принимается за основной тон, а нота, соответствующая данной частоте считается нажатой, а все ее гармоники исключаются из дальнейшей проверки. Если же какая-либо из гармоник этой ноты громче основного тона (а по условию, все они тише), значит, имеет место наложение гармоник данной ноты на частоты другой, и потому рекурсивно вызывается проверка этой гармоники (которая оказалась громче основного тона). Гармоника проверяется точно так же, как если бы проверялась любая другая нота, у которой тоже, в свою очередь, могут быть свои гармоники. После завершения этой рекурсивной проверки алгоритм продолжает свою итеративную работу. Важно отметить, что



работа алгоритма начинается по возрастанию частот, поскольку, как уже было отмечено, при успешной проверке, гармоники зажатых нот исключаются из проверки и в дальнейшем не будут приняты по ошибке за основные тона. Детектированная нота считается отпущенной, когда в последующие моменты времени ее амплитуда совершит значительный резкий скачок, либо станет приблизительно равной нулю.

Здесь стоит также отметить еще одну трудность распознавания. Так как частоты нот низких октав находятся очень близко друг к другу, например, с разницей в 5 Гц (ноты D#2(77Гц) и D2(73Гц)), а шаг частотной сетки FFT равен, например, 11.72 Гц (как из примера выше) то амплитуды этих составляющих сольются в одну и будут иметь место некоторые погрешности.

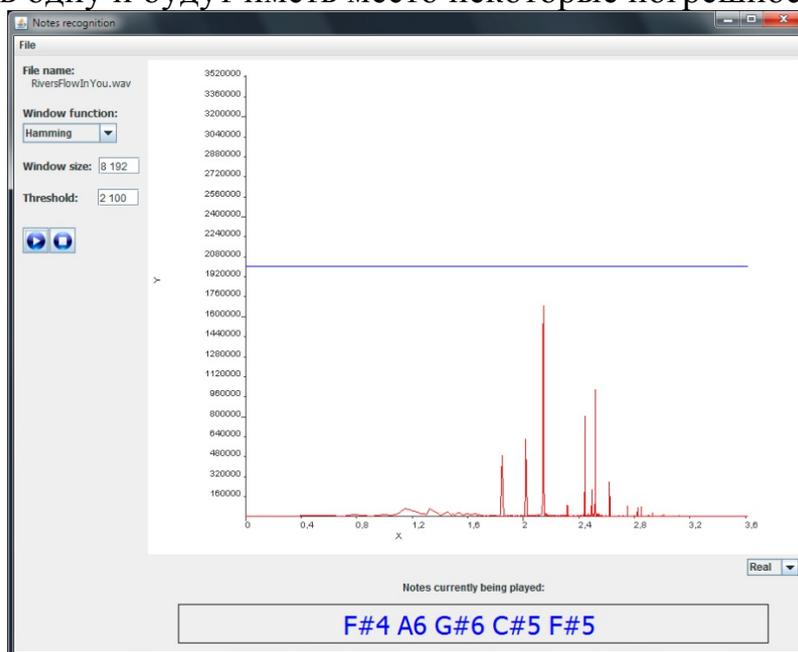


Рис. 4. Интерфейс программы

### Выводы

Разработанному алгоритму с некоторой точностью удается распознавать ноты фортепиано выше первой октавы. Требуется доработка для распознавания нот всех октав, что возможно будет включать в себя применение нейронных сетей, а также применение более точных алгоритмов взятия спектра сигнала для поставленной задачи.

### Литература

1. Сайт ProSound [Электронный ресурс]: статьи. – Электрон. дан. – Режим доступа: <http://prosound.ixbt.com/> – Загл. с экрана. – Яз. рус.
2. Википедия [Электронный ресурс]: статьи. – Электрон. дан. – Режим доступа: <http://ru.wikipedia.org/> – Загл. с экрана. – Яз. рус.
3. Хабрахабр [Электронный ресурс]: статьи. – Электрон. дан. – Режим доступа: <http://habrahabr.ru> – Загл. с экрана. – Яз. рус.