Реализация и анализ алгоритмов распознавания высоты звуков в музыкальных фрагментах

Н. В. Воинов[1], Д. А. Иванов, С. А. Молодяков, Т. В. Леонтьева
Санкт-Петербургский политехнический университет Петра Великого
[1]voinov@ics2.ecd.spbstu.ru

***Аннотация*. Работа посвящена исследованию алгоритмов распознавания высоты звуков в музыкальных фрагментах. Проведен обзор наиболее распространенных алгоритмов распознавания нот в одноголосных мелодиях (single-pitch estimation), представлена их программная реализация, на основании полученных результатов работы алгоритмов проанализированы их точность, эффективность и применимость к решению задачи из области multi-pitch estimation – распознавания аккордов. Для устранения выявленных недостатков существующих алгоритмов разработан собственный алгоритм на основе полносвязной трехслойной нейронной сети для распознавания аккордов в музыкальных фрагментах, оценены результаты его работы.**

Ключевые слова: распознавание высоты звука; нейронная сеть; одноголосные мелодии; распознавание аккордов; цифровая обработка сигналов

I. Введение

Существующие технологии и алгоритмы цифровой обработки сигналов способствуют автоматизации задачи распознавания высоты звуков. Автоматическое распознавание высоты звуков – трудоемкий и сильно зависимый от исходных данных процесс: звуковые файлы, подлежащие распознаванию, должны быть корректно раскадрированы, к ним должна быть применена подходящая оконная функция, нежелательно присутствие шумов. Возникают определенные сложности и непосредственно в процессе распознавания: громкость и тембр инструмента, на котором воспроизводится мелодия, а также специфичные приемы игры, реализованные на нем, могут негативным образом повлиять на результат распознавания.

Задача определения высоты звуков в одноголосных мелодиях (single-pitch estimation) позиционируется как уже решённая. В данной работе проводится исследование точности и эффективности распознавания уже известных алгоритмов, специализирующихся на single-pitch estimation, именно в работе с данными, которые могут вызывать сложность в распознавании. Под точностью здесь понимается, насколько близко значение распознанной частоты звукового сигнала располагается по отношению к эталонной частоте (при правильной настройке инструмента). Эффективность алгоритмов в подавляющем большинстве случаев заключается в их способности фильтровать шумы, фиксировать паузы в музыкальных фрагментах и корректно интерпретировать фрагменты с приемами игры на музыкальных инструментах.

Смежной является задача из области multi-pitch estimation – распознавание аккордов. В настоящий момент точность результатов при автоматическом определении аккордов не превышает 70 % [1, 2], что делает работы в данном направлении весьма актуальными и востребованными. В данной работе представлен разработанный алгоритм по распознаванию аккордов, основанный на реализации нейронной сети. Исследованы точность и эффективность алгоритма при работе с музыкальными фрагментами.

Применение нейронной сети обуславливается недостаточной эффективностью существующих алгоритмов single-pitch estimation в задаче распознавания аккордов, что было установлено в ходе их анализа.

II. Анализ Алгоритмов Single-Pitch Estimation

Ранние алгоритмы по оценке высоты звука осуществляли анализ непосредственно по форме входящего звукового сигнала. Работа Бернарда Голда [3] была своего рода одной из первых попыток создания надёжной системы по определению высоты звуков. Система, описанная в работе, была основана на нахождении шаблонов и закономерностей в необработанном цифровом представлении звукового сигнала, а конкретно на нахождении пиков амплитуды и повторяемости этих пиков. Так, идея нахождения периодичности в исходных цифровых представлениях звуковых сигналов послужила основой для целого семейства алгоритмов, базирующихся на вычислении корреляции. В работе был проведен анализ работы двух значимых алгоритмов этого семейства – функции автокорреляции, описанной Лоуренсом Рабинером [4], и связанной с ней нормализованной кросс-корреляционной функции. Оба алгоритма сканируют форму волны входного сигнала, генерируют набор весов, соответствующих предположениям о периоде сигнала, выбирают наибольшее значение и подают его на выход системы. Корреляционные функции хороши своей резистентностью к шумам и надёжностью. Однако у них есть и два существенных недостатка: возникновение вероятности нахождения кратной периодичности в форме сигнала и требование к сравнительно большой длительности самого сигнала [5].

Одной из аудиозаписей, на основе которой производился анализ результатов работы алгоритмов, послужила запись довольно сложной скачкообразной мелодии, сыгранной на трубе. Результаты распознавания

данной мелодии в частотном и нотном выражении показаны на рис. 1 и рис. 2 соответственно. В самом начале можно зафиксировать резкий необычный скачок на октаву вниз. Данная ошибка в распознавании обусловлена спецификой духовых инструментов. Также можно заметить, что последняя очень краткая по звучанию нота не была распознана вообще, что обусловлено требованием алгоритма к более длинной продолжительности нот для корректного распознавания, о чём упоминалось ранее.

Самый существенный недостаток данного семейства алгоритмов заключается в том, что алгоритмы подразумевают работу с единичными квазипериодическими цифровыми представлениями звуковых волн. Это объясняется следующим: если форма волны звукового сигнала составлена двумя музыкальными нотами, периодичными по отдельности в интервалах T1 и T2, то результирующее представление звукового сигнала не будет периодично ни на T1, ни на T2, если только одна величина не кратна другой. Именно этот факт не позволяет использовать данные алгоритмы в контексте задач определения высоты нескольких звуков в один момент времени.

Наряду с методами определения высоты звуков через период звуковых сигналов существуют методы, работающие и со спектром сигнала. Спектр извлекается с использованием быстрого преобразования Фурье для каждого кадра исходной аудиозаписи. Вместо попыток сканирования формы исходного звукового сигнала алгоритмы работают с самим спектром сигнала после преобразования. В процессе работы со спектром главная задача определения высоты звука заключается в правильной интерпретации закономерности повторения частот в спектрограмме. Спектр отдельно взятой ноты сам по себе является периодическим, где базовая частота и является периодом. Таким образом, все последующие в спектре частотные гармоники расположены равномерно через значение базовой частоты. Все алгоритмы, работающие со спектром звуковых сигналов, так или иначе используют данный принцип повторяемости частотных гармоник. Серьёзной задачей является непосредственно само определение базовой частоты. Данный процесс может быть затруднён из-за наличия шумов в исходной аудиозаписи или из-за специфики тембра инструмента. Из множества таких алгоритмов, работающих со спектром сигнала, для анализа был выбран кепстральный алгоритм и алгоритм, базирующийся на методе частотных гармоник.

По результатам программной реализации перечисленных выше алгоритмов и анализа их работы были сделаны выводы об их характеристиках и применимости к решению задачи распознавания аккордов (таблица).

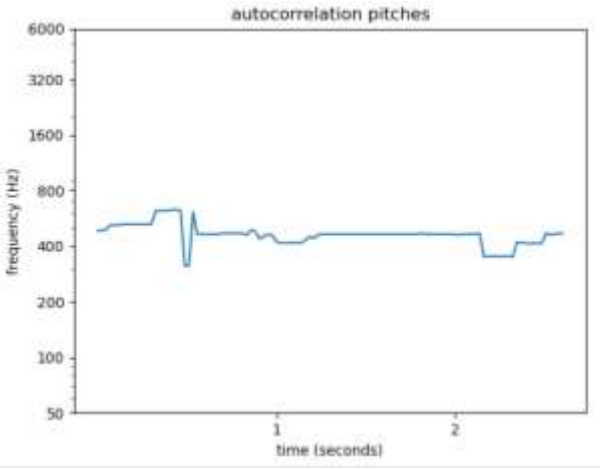

Рис. 1. График распознанных частот из записи трубы для автокорреляционной функции

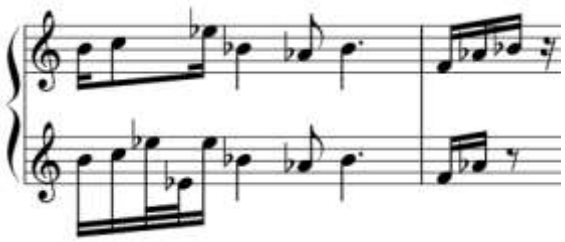

Рис. 2. Эталонная мелодия трубы (сверху) и нотное представление результатов распознавания (снизу)

В таблице алгоритмы обозначены следующим образом: 1 – автокорреляция, 2 – кросс-корреляция, 3 – кепстральный алгоритм, 4 – метод частотных гармоник.

ТАБЛИЦА I СРАВНЕНИЕ АЛГОРИТМОВ SINGLE-PITCH ESTIMATION

Критерии	**Алгоритмы**			
	1	***2***	***3***	***4***
Игнорирование шумовых частот	Высокое	Низкое	Среднее	Высокое
Игнорирование дублирующих октавных частот	Среднее	Среднее	Высокое	Высокое
Распознавание пауз	Нет	Нет	Нет	Да
Распознавание нот при звучании других нот на фоне	Да	Нет	Да	Да
Потеря звучащих нот	Средняя	Низкая	Нет	Нет
Частотная точность	Средняя	Высокая	Очень Высокая	Очень высокая
Общая эффективность	Средняя	Низкая	Высокая	Высокая
Универсальность	Низкая	Нет	Средняя	Средняя
Применимость для multi-pitch estimation	Низкая	Низкая	Средняя	Средняя

III. Разработка алгоритма для распознавания Аккордов

Необходимость разработки алгоритма, основанного на нейронной сети, для распознавания аккордов в музыкальных фрагментах обуславливается следующими факторами:

- недостаточная эффективность алгоритмов single-pitch estimation для задач multi-pitch estimation;
- требование универсальности к разрабатываемому решению, что выполняется благодаря применению нейронной сети.

Важно отметить, что алгоритмом решается задача классификации аккордов. В такой классификации будет 10 классов в соответствии с количеством наиболее часто встречающихся аккордов (A, Am, Bm, C, D, Dm, E, Em, F, G) [6].

Входные данные представляются следующим образом. Используется быстрое преобразования Фурье и Constant-Q преобразование [7], которое интерпретирует частоты в логарифмическом масштабе для отображения звуковых частот в контексте их различимости человеческим ухом. После этих двух преобразований звуковой сигнал можно представить в виде вектора длиной в 12, именуемого "профилем высотного класса" (PCP – Pitch Class Profile) [8]. Длина обуславливается количеством нот в одной октаве. Таким образом, при конструкции вектора [G, G#, A, A#, B, C, C#, D, D#, E, F, F#] трезвучие "До-мажор" (C/C-dur: C + E + G) в "вакууме" будет представляться таким вектором как [1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0]. Однако на практике невозможно встретить такой бинарный вектор: на записи, подлежащей преобразованиям, всегда будут присутствовать шумы, отзвуки и обертоны, которые также влияют на значения вектора. Данное представление не зависит ни от расположения аккорда (тесного или широкого), ни от октавы, в которой он сыгран: данный вектор предоставляет информацию только о присутствующих в аккорде нотах вне контекста октавной высотности. Таким образом, на вход нейронной сети будет подаваться вектор из 12 значений.

Для реализации алгоритма была сконструирована полносвязная трёхслойная нейронная сеть со слоями размерности 12, 10 и 10 соответственно. Схема нейронной сети изображена на рис. 3. Полная схема алгоритма представлена рис. 4.

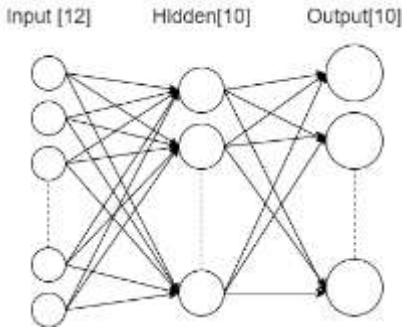

Рис. 3. Схема нейронной сети

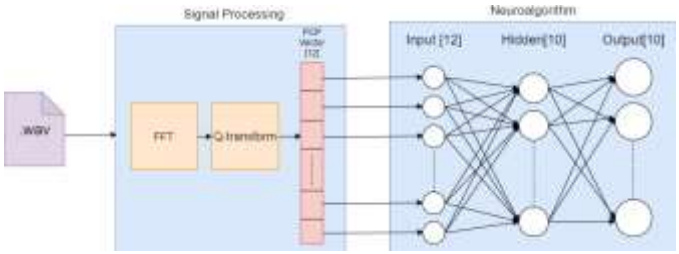

Рис. 4. Схема алгоритма

IV. Результаты

Результаты обучения нейронной сети представлены на рис. 5. Видно, что точность классификации стремится к абсолютной, что является отличным результатом. При этом отсутствует эффект переобучения: потеря в распознавании на «validation»-выборке в соответствии с рис. 6 с каждой эпохой уменьшается.

Обучение стоит продолжать, пока показатель ошибки уменьшается. Возможно, при дальнейшем увеличении количества эпох он также будет уменьшаться, но к шестидесятой эпохе уже достигается стопроцентная точность классификации, поэтому дальнейшее увеличение количества эпох не рассматривалось.

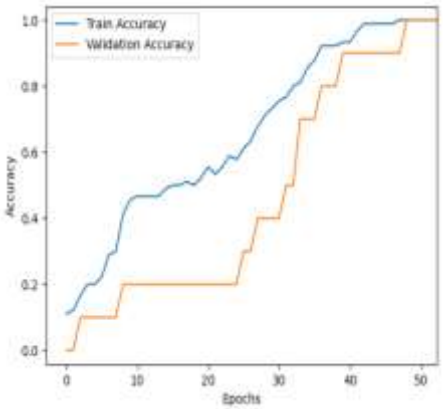

Рис. 5. Точность классификации разработанного алгоритма

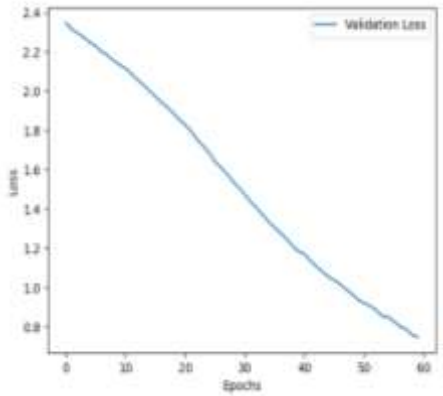

Рис. 6. Ошибка на «validation»-выборке

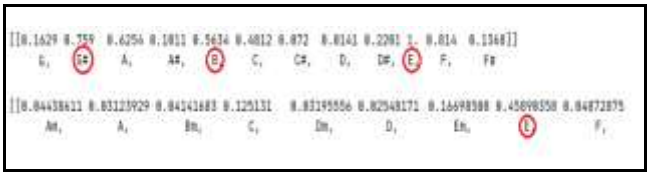

Рис. 7. Результаты классификации аккорда

Сформируем предсказание об известном заранее аккорде E (E – трезвучие «ми-мажор»: ноты «ми», «соль-диез», «си») в реализованной системе. Для этого подадим сформированный PCP-вектор этого аккорда для предсказания. Результаты предсказания отражены на рис.7. Наибольшее значение (0,4589) в выходном векторе классов имеет класс, соответствующий аккорду E-аккорду, который и подавался на самом деле на вход системе. Примечательно, что вторым (0,1669) и третьим (0,1251) предсказанием в соответствии с выходными значениями системы являются аккорды Em и Cm соответственно. Объясняется это довольно просто: эти два аккорда по своему нотному составу очень похожи на аккорд E. Тем не менее, система выдаёт предположение об аккорде E с весом почти в три раза больше второго по значению предположения.

V. Заключение

В работе проведен обзор наиболее распространенных алгоритмов распознавания нот в одноголосных мелодиях (single-pitch estimation), на основании полученных результатов работы алгоритмов проанализированы их точность, эффективность и применимость к решению задачи из области multi-pitch estimation – распознавания аккордов. Для устранения выявленных недостатков существующих алгоритмов разработан собственный алгоритм на основе полносвязной трехслойной нейронной сети. Точность распознавания алгоритма достигает 95–99 % на тестовой выборке. При этом можно говорить о высокой универсальности алгоритма вследствие разнообразной по тембру и количеству шумовых эффектов и отзвуков коллекции аудиозаписей, на основе которых производилось обучение.

Тем не менее, в работе алгоритма есть ряд ограничений, преодоление которых ещё более увеличит универсальность распознавания аккордов. Главным ограничением является отсутствие точной структуры аккорда и точного октавного расположения всех входящих в него нот при получении результата. Это ограничение обуславливается форматом входных данных, который не предусматривает информацию о таком октавном расположении нот. Также существенным ограничением является малое количество аккордов, доступных для классификации, так как в процессе реализации было выбрано всего 10 самых распространённых аккордов. Данное ограничение можно преодолеть, если расширить количество классов в системе, предварительно записав для них определённое количество разнообразных wav-файлов, которые будут служить дополнительным набором обучающих данных для системы.

Список литературы

[1] O'Brien C., Plumbley M.D. Automatic music transcription using low rank non-negative matrix decomposition. 25th European Signal Processing Conference, 2017, vol. 2017-January, pp. 1848-1852. DOI: 10.23919/EUSIPCO.2017.8081529

[2] Multiple Fundamental Frequency Estimation & Tracking Results - MIREX Dataset, 2017: https://www.music-ir.org/mirex/wiki/2017:Multiple_Fundamental_Frequency_Estimation_%26_Tracking_Results_-_MIREX_Dataset (дата обращения 09.03.2021)

[3] Gold B. Computer Program for Pitch Extraction. Journal of the Acoustical Society of America. 1962, vol. 34, no. 7, pp. 916-921. DOI: 10.1121/1.1918221

[4] Rabiner L.R. On the Use of Autocorrelation Analysis for Pitch Detection. IEEE Transactions on Acoustics, Speech, and Signal Processing. 1977, vol. 25, no. 1, pp. 24-33. DOI: 10.1109/TASSP.1977.1162905

[5] Talkin D. A robust algorithm for pitch tracking (RAPT). Speech Coding and Synthesis. 1995, pp. 495-518.

[6] Carlton D. I analyzed the chords of 1300 popular songs for patterns. This is what I found: https://www.hooktheory.com/blog/i-analyzed-the-chords-of-1300-popular-songs-for-patterns-this-is-what-i-found (дата обращения 09.03.2021)

[7] Schörkhuber C., Klapuri A. Constant-Q transform toolbox for music processing. Proceedings of the 7th Sound and Music Computing Conference. 2010, p. 20.

[8] Fujishima T. Realtime Chord Recognition of Musical Sound: a System Using Common Lisp Music. ICMC Proceedings. 1999, pp. 464-467.