

МУЗЫКАЛЬНЫЙ ИНФОРМАЦИОННЫЙ ПОИСК С ЗАПРОСОМ ПО НАПЕВУ

Безрукова Е.М.

(г. Томск, Томский Политехнический Университет)

QUERY BY HUMMING IN MUSIC INFORMATION RETRIEVAL

Bezrukova E.M.

(Tomsk, Tomsk Polytechnic University)

In this paper the problems of Query by Humming in Music Information Retrieval systems are analyzed. A statistical approach to the problem of retrieval is presented. The processes of segmentation as well as of the extraction of pitch and duration data are described. From the extracted data the characteristic vector is formed for each segment. The method of using the vectors in melodic search is proposed.

Введение. Музыкальный информационный поиск, или МИП – междисциплинарное научное направление, использующее различные методы обнаружения знаний (data mining) в данных, представляющих музыкальные отрывки. Круг задач МИП довольно широк [1], однако в рамках данного исследования следует ограничиться следующими направлениями:

1) Предобработка, куда входят задачи эффективного кодирования, декодирования, сжатия, сегментации, подавления шумов, распознавания голоса и отдельных инструментов;

2) Извлечение определенных музыкальных характеристик: ритма, гармонии, тональности, мелодической составляющей и т.п.;

3) Собственно поиск, куда входит определение и обработка субъективных характеристик музыки (жанр) и объективных (тональность, ритм, гармония, мелодия, акустические характеристики, метаданные об исполнителе, названии, альбоме и т.п.). Сюда же входят задачи обработки запроса по образцу.

Запрос по образцу – метод поиска музыки на основе предоставленного образца. В качестве образца может выступать записанный отрывок искомого музыкального произведения. Однако чаще всего возникает необходимость в нахождении музыкального произведения только на основе запомненного отрывка. В этом случае пользователь может напеть, наиграть, насвистеть мелодию или попробовать представить ее по памяти в символьном виде, с помощью классической нотации или других видов кодирования, например, в коде Парсонса.

Напевание – наиболее естественная для человека форма воспроизведения музыки. Однако такая форма представляет определенные трудности для обработки, т.к. при воспроизведении музыки голосом и по памяти возможны неточности [2]:

- Неверная тональность;
- Неверные интервалы в мелодии;
- Неверный ритм или длительность нот.

Кроме того, на качество распознавания мелодии влияет качество записи, которая может быть сильно зашумлена или искажена.

Предметом данной работы является исследование способа обнаружения мелодии в звуковом отрывке с записью голоса человека (далее – диктор), напевающего мелодию. Полифонические музыкальные отрывки не анализируются.

Предлагаемый подход. Существует несколько подходов к решению задачи музыкального информационного поиска с запросом по напеву [2]. В данной работе предлагается следовать аналогии с процессом восприятия мелодии у человека.

Человек воспринимает мелодию как информацию об относительной высоте тона и продолжительности звучания нот, которые выстраиваются в мелодическую последовательность [3]. В классической нотации высоте тона присваивается абсолютное значение, например, 440 Гц соответствует ноте ля первой октавы. Однако применительно к рассматриваемой задаче, где мелодия напевается голосом, в использовании информации об абсолютной высоте тона нет необходимости. Диктор может начать напевать с неверной ноты и в неверной тональности, и, тем не менее, мелодический рисунок остается узнаваемым. Такое явление называется транспонированием. При использовании в алгоритме информации об изменении высоты тона относительно первой или предыдущей ноты проблема транспонирования в неверную тональность решается автоматически.

Подобным же образом аргументирован отказ от использования в алгоритме абсолютных значений длительности нот (четвертей, восьмых и т.п.) в пользу относительных.

Основные этапы алгоритма. Первым шагом для распознавания мелодии является сегментация записанного звукового отрывка. Учитывая относительный характер длительности нот, было решено не разбивать мелодию на участки одинаковой длительности, а воспользоваться принципом динамического разбиения на слоги, лежащие в основе систем распознавания речи и использующие скрытые Марковские модели (СММ) [4]. Границы между сегментами определяются изменением интенсивности, высоты и спектра, при этом необходимым условием является напевание мелодии слогами «на-на-на», «ла-ла-ла» и т.п.

После разбиения на сегменты необходимо определить высоту тона для каждого сегмента. Для этого применяется стандартное преобразование Фурье [4].

Далее необходимо определить изменения по высоте и длительности звучания соседних нот. Относительное изменение высоты тона (ОИВ) определяется по следующей формуле:

$$\text{ОИВ} = \frac{\log(f_k) - \log(f_{k-1})}{\text{ТКП}}$$

где f – частота звука, k - порядковый номер звука в мелодии, ТКП - теоритический коэффициент распределения полутонов, $\log^{12}\sqrt{2}$. Относительное изменение длительности (ОИД) ноты определяется по формуле:

$$\text{ОИД} = \frac{t_k}{t_{k-1}}$$

где t – продолжительность звучания k -й ноты.

Таким образом, сегмент мелодии представляется характеристическим вектором [Двысота, Двремя], описывающий изменение высоты и длительности по сравнению с предыдущим сегментом. Последовательность пар изменений этих характеристик для всех сегментов представляет мелодию. Конечным этапом музыкального информационного

поиска является сравнение двумерных характеристических массивов напетого диктором отрывка с имеющимися в базе данных массивами и нахождение наиболее близкого по значениям элемента в базе.

Заключение. В данной работе исследуются основные проблемы информационного музыкального поиска с запросом по напеву, а также предлагается метод, устойчивый к одной из них и сглаживающий влияние других. Неосвещенным остался вопрос подавления шумов на записи, т.к. для этого существует множество готовых решений. Кроме того, пока не решен вопрос поиска наиболее похожего музыкального отрывка в базе данных, который является основным предметом исследований автора на данный момент.

ЛИТЕРАТУРА

1. Fingerhut M. Music Information Retrieval, or how to search for (and maybe find) music and do away with incipits // IAML - IASA 2004 Congress, Oslo. - 2004
2. Weinstein E. Query by Humming: a Survey. - 2006
3. Unal E., Narayanan S.S., Chew E. A Statistical Approach to Retrieval under User-dependent Uncertainty in Query-by-Humming Systems. // MIR'04, October 15- 16, 2004
4. Shih H.-H., Narayanan S.S., Jay Kuo C.-C. Multidimensional Humming Transcription Using Hidden Markov Models for Query by Humming Systems. - 2003

КОНЦЕПТУАЛИЗАТОР ГРАФИЧЕСКОГО ИНТЕРФЕЙСА ПОЛЬЗОВАТЕЛЯ С РАЗНОРОДНЫМИ ИСТОЧНИКАМИ ДАННЫХ БОЛЬШОГО ОБЪЕМА

*Л. Ю. Исмаилова, А. А. Борзяк, А. Н. Долбин, А. С. Доронин, М. Ю. Ермак,
С. В. Косиков, М. А. Маслов, В. В. Навроцкий, В. Н. Назаров, М. Л. Файбисович
(г. Москва, ГК "ЮрИнфоР"/НИЯУ МИФИ)*

CONCEPTUALIZATOR FOR THE GRAPHICAL USER INTERFACE TO HETEROGENEOUS BIG DATA SOURCES

*L.Yu. Ismailova, A. A. Borzyak, A. N. Dolbin, A. S. Doronin, M. Yu. Ermak,
S. V. Kosikov, M. A. Masolv, V. V. Navrotskiy, V. N. Nazarov, M. L. Faybisovich
(с. Moscow, GK "JurInfoR/NRNU MEPhI")*

It is proposed a conceptual prototype of a graphical user interface (CGI) to user databases. The system is aimed primarily at non-professional users in the IT field and allows them to obtain information from heterogeneous data sources, including object-oriented and XML-oriented representations. The system is newly developed by the expansion of the original module Active Extensional (AcE), which will be used as an intermediary between the data and the user.

Введение. В настоящее время наблюдается особенно быстрый рост массы цифровых данных, а по данным INTEL около 90% имеющихся данных появилось за последние 2 года. Критическими факторами при работе с данными являются их объем, скорость передачи/преобразования, разнообразие и ценность. В этих условиях для обеспечения технологически эффективной цепочки хранения-обработки-доступа к