Yevheniia Kolomiiets, Polina Shakhova, Artem Volokyta

AUDIO FEATURES EXTRACTION FOR NEURAL NETWORKS USAGE

The article deals with the issue of feature extraction of an audio signal for using the result data by a neural network that identifies the elements of harmony of a musical piece. The developed system is an audio data preprocessing service and uses such algorithms as the tonal profile algorithm (KSH algorithm) and Python, Numpy, and Librosa language tools to determine key audio.

Key words: audio data, audio analysis, music analysis, Python, audio processing.

Fig.: 4. Bibl.: 5.

Relevance of the research topic. Audio analysing services are a popular area of software development because of the demand in tools like audio identification, classifying audio and extraction of audio characteristics [1] in audio streaming and creation services for improving their algorithms (such as recommendations, storing and editing).

Tools for identifying audio characteristics are also widely used by practicing musicians.

Target setting. Due to the relevance of the subject of audio analysis, there are multiple services that deal with different aspects of the feature extraction, the difference being that this research focuses on preprocessing algorithms for normalizing audio signal values.

Actual scientific researches and issues analysis. Audio features extraction presents a complex problem and as a result has an entire field dedicated to evaluating and improving the findings of different audio processing systems.

One of the persistent research practices are present as The Music Information Retrieval Evaluation eXchange (MIREX). [2] This is a state-of-the-art, research-based approach to music analysis coordinated and managed by the International Music Information Retrieval Evaluation Laboratory. There are many different methods used by researchers being used and evaluated in the field of music data retrieval (MIR).

Some of the researched methods and functions, such as Sound Onset Detection, are small-scale MIR detection (e.g., identifying the locations of music starting points in audio files that match the index). Others, such as Symbolic

Melodic Similarity, are the MIR studies that operate at a high level (e.g. the creation of music based on patterns of similarity).

Uninvestigated parts of general matters defining. As a result of reviewing the existing audio analysis systems and comparing the components related to low-level data about musical pieces, the inefficiency of these algorithms according to the preprocessing aspect was found. Some of the systems have a complete set of methods for obtaining data about works, but it are limited critically – users can only get the pre-calculated information about audio that is present in the service's database.

Some systems on the other hand support analyzing and loading of user data but do not return data in the required format - only the result of calculations and are often critically inaccurate when it comes to individual values.

The research objective. This article has objective of researching music and signal parameters that are useful for extracting harmonic audio components, [1] developing modules that calculate these features and creating a service for audio preprocessing that allows to split into segments and calculate tonal characteristics of an audio and to save this data for usage by chord recognizing neural network.

The statement of basic materials. The system operates on such concepts and algorithms as reading the audio in WAV format, performing a preliminary analysis and determining the core characteristics of the piece – key and BPM, which will be saved for other stages, using the algorithm of tonal profiles and Numpy tools for the key identification and onset detection functions of Librosa [4] for the latter purpose.

The main component of work is defining the beat of the music and forming frames so that one frame represents the duration of the audio from one beat to the next and calculating the tonal features (in this case MFCC or Mel spectrogram) for each of the frames separately.

General system structure. The developed system should contain all the components needed to extract audio features and to perform normalization and storage of the extracted data in a file in order to save it on disk for further operations.

The chosen way perform this task is illustrated on the figure 1. The diagram illustrates the concepts of working on the segments of the audio separately and treating the subsystem of chord identification as a “black box”, where only preservation of tone-rhythm values relation is important.

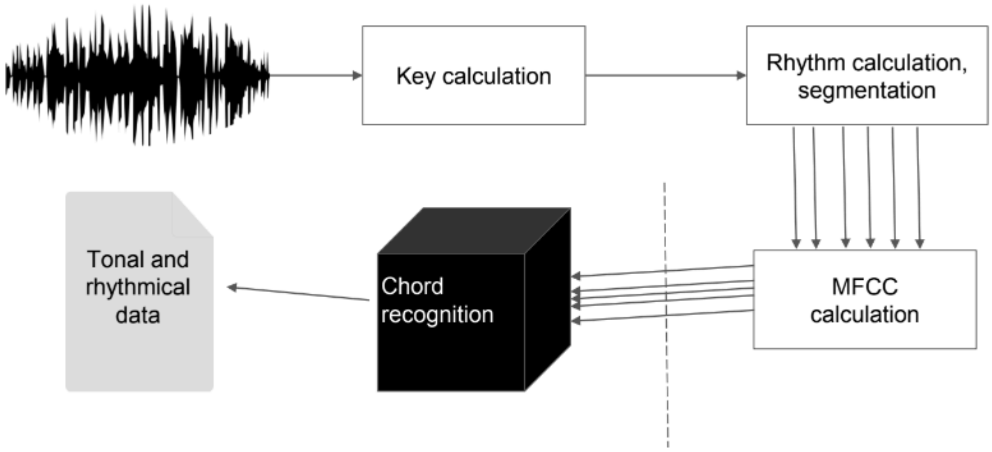

Figure 1. General system components

Key calculation. For key calculation the tonal profiles algorithm is used, another name being the Krumhansl-Schmuckler algorithm. [5] For effective array operations NumPy is used.

The main steps of the algorithm can be distinguished as:

1. Receiving the input data.
2. Calculating the tonal representation – chroma features from input data.
3. Calculating the sums of the values of each of the 12 notes over the duration of the entire audio – thus obtaining an audio profile.
4. Finding the correlation coefficients between the obtained audio profile and each of the 12 tonality profiles, more specifically between the theoretical tonality profile and the profile of the studied audio.
5. From the received coefficients, choosing the largest value - it corresponds to the most likely tonality result. It is also possible to obtain an alternative tonality corresponding to the second highest coefficient.
6. Returning the received values.

The flowchart representation is illustrated at figure 2.

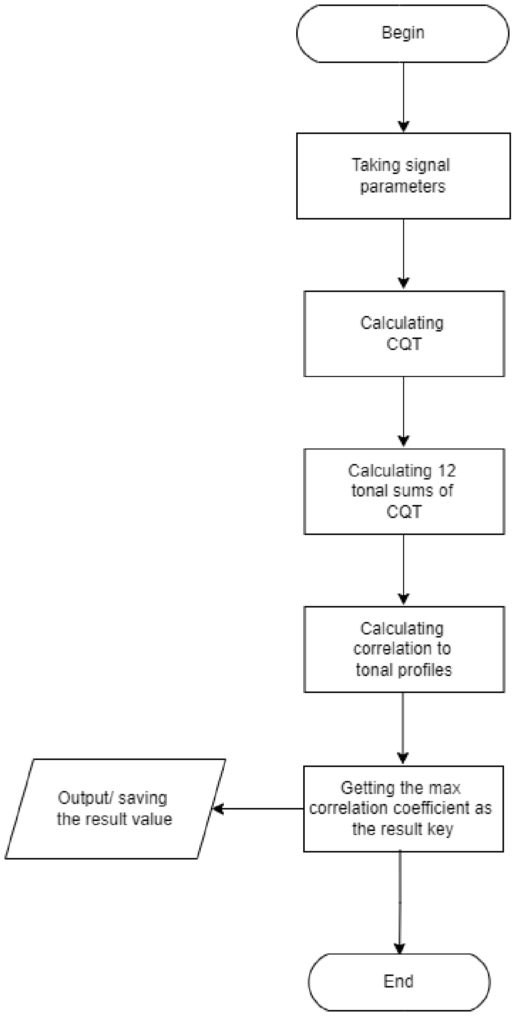

Figure 2. Flowchart of key detection component

Audio features. Different spectral features represent different relations. [4] Wave plot demonstrates the time-amplitude relation, chroma feature illustrates the tone-time values and STFT shows the relation between absolute frequency and time (figure 3).

The chosen formats of audio spectral features are MFCC та Mel spectrogram, which represent frequency-time relation with the help of mel cepstral coefficients. They are different from STFT because they are preserving timbre, don’t store absolute frequency, but store value according to the human ear, and are the most popular features for training neural networks – more details in related work.

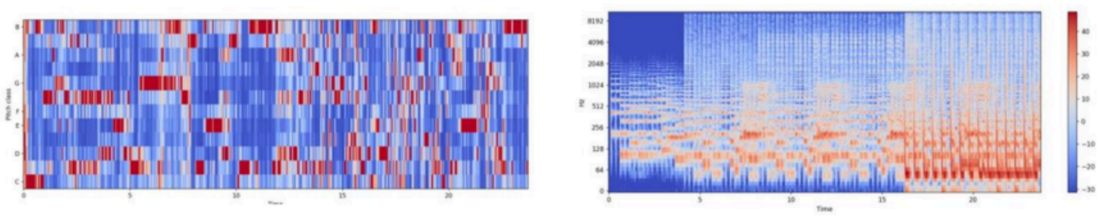

Figure 3. Spectral features of audio signal

Evaluation of the results. Algorithm of rhythm calculation was tested for the results correctness: accuracy of results is preserved up to 0.03 deviation, where Spotify and visualisation was used for theoretical values.

Algorithm of key calculation was also tested for the results correctness: inaccurate results could be observed in songs with weak harmonical components with average accuracy being 0.8. Spotify [3] and musical theory were used for obtaining theoretical values.

Files with the results data are presented as JSON files with metadata header, where one of the fields has segment data, presented as headers and MFCC features (figure 4).

```
{
    "name": "Alejandro Gaga Lady",
    "BPM": 99.38401442307692,
    "key": "G",
    "alt_key": "D",
    "frame_size": 512,
    "sample_rate": 22050,
    "mfccs": [
        {
            "header": {
                "id": 0,
                "frame_start": 0,
                "frame_end": 586
            },
            "mfcc": [
                [
                    -715.8488159179688,
                    0.500795304775238,
                    0.5006826519966125,
                    0.5004936456680298,
                    0.5002302527427673,
                    0.4998905062675476,
                    0.49947649240493774,
                    0.49898630380630493,
                    0.49841994047164917,
```


Figure 4. Results data file

A tool for data extraction was also developed, which allows to get WAV audio file with a YouTube URL. Tool is working correctly, but some predictable loss of quality is observed compared to pure WAV file.

Conclusions. As a result of research, the following conclusions were made:

1. The audio parameters were researched and it has been found that key, rhythm and tonal features should be extracted for proper audio evaluation and neural network teaching

2. The tools for audio visualization and feature calculation were developed, and it has been affirmed that the developed system works correctly and has an acceptable accuracy of results

3. The format of the result data was found to be readable and convenient.

As the results of research and evaluation it is safe to say that the results could be useful for appliance in the field of Musical Information Retrieval. The next step of music data usage is the neural network for chord recognition.

Some possible improvements could be suggested for future work: adding the ability to identify the work not by name and regular expression, but by audio fingerprint and implementing of working with more audio formats.

References

1. Phillip Magnuson. A Structural Examination of Tonality, Vocabulary, Texture, Sonorities, and Time Organization in Western Art Music [Electronic resource] / Phillip Magnuson – Mode of access to the resource: https://academic.udayton.edu/phillipmagnuson/soundpatterns/compbasics/5otherparts.html.
2. The Music Information Retrieval Evaluation eXchange (MIREX) [Electronic resource] – Mode of access to the resource: https://www.dlib.org/dlib/december06/downie/12downie.html
3. Spotify. Web API Guides [Electronic resource] / Spotify – Mode of access to the resource: https://developer.spotify.com/documentation/web-api/guides/.
4. Mcfree B. librosa: Audio and Music Signal Analysis in Python / Mcfree Brian, 2015. – p. 24.
5. Madsen S. KEY-FINDING WITH INTERVAL PROFILES / Madsen Søren – Linz. – p. 4.

AUTHORS

Kolomiiets Yevheniia – bachelor, Department of Computer Engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.
E-mail: jennichka10155@gmail.com

Shakhova Polina – bachelor, Department of Computer Engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.

E-mail: schachowa.paulina@gmail.com

Volokyta Artem – associate professor, Department of Computer Engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.
E-mail: artem.volokita@kpi.ua