# MemoVision

Link to the software: https://www.memovision.phil.muni.cz/

Link to the GitHub repo: https://github.com/stepanmk/memovision

**Idea and motivation:**

This document introduces the MemoVision software. The idea behind MemoVision software is to analyze multiple interpretations (or performances) of a musical piece and compare their differences (usually classical music thanks to its interpretation variance while keeping the overall structure) and provide user interface to analyze the data. The analysis is focused on objective parameters that can distinguish how performers play the piece (how they differ in tempo, dynamics, expressivity).

The software provides easy-to-use graphical interface with prebuild functions to upload recordings (from your local PC), additional metadata, labels, and reference measures. The reference measures are a vector of values corresponding to the time position of measures (bars) of a given "reference" recording. It needs to be in a correct format: one column of time values or two columns (time value and beat position). For instance, see Figure 1. Figure 2 shows the measure positions for an audio waveform in Sonic Visualiser (software for, e.g., creating the ground-truth measure positions). The reference recording can be any recording as long as the audio quality, tempo deviations, and especially structural integrity (following the same sheet music or score) are comparable to all other performances. In that case, the synchronization between recordings should be successful and analysis possible.

```
0.2716099770
1.7360544220
3.0042176870
4.2408163270
5.5338775510
6.8355102040
8.4983673470
9.9282993200
11.3824489800
13.2265306120
14.5648979590
```

Fig. 1: Text file (.txt or .csv) with time positions of measures for a reference recording.

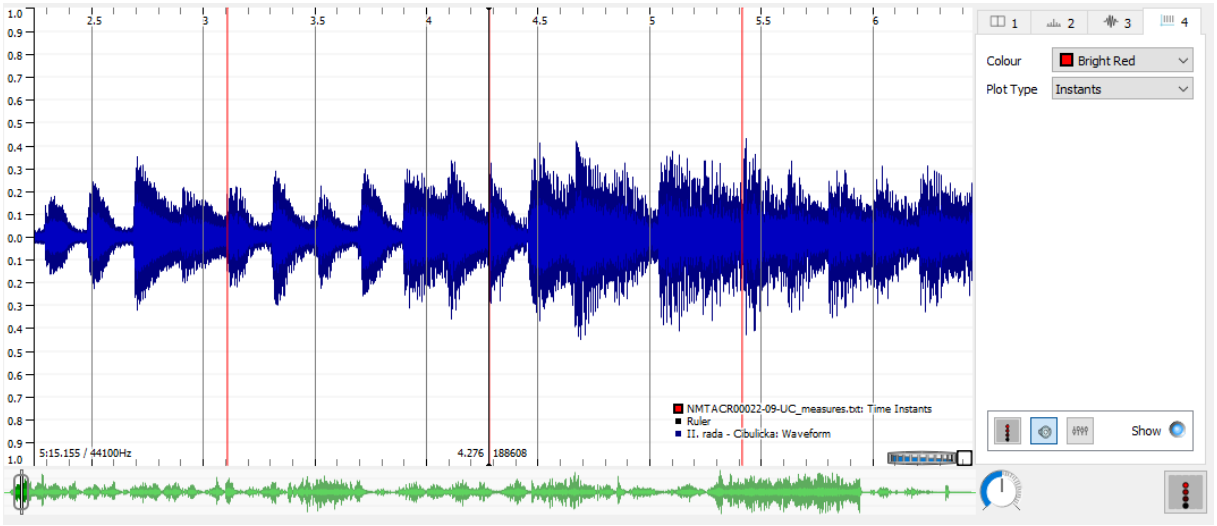


Fig. 2: Time positions of measures for a reference recording in Sonic Visualiser.

The main concept is to use audio-to-audio synchronization to transfer the reference measure positions to all other versions (performances). This is done using MrMsDTW (memory-restricted multiscale DTW) algorithm with beat tracking neural network to increase the measure-level synchronization accuracy. We trained the TCN beat tracking network to follow the synctoolbox pipeline. The input recordings are sampled at 22,05 kHz and the output function has a resolution of 50 fps (same as chroma vectors). By synchronizing all recordings, we transfer the measure positions by interpolation via final alignment path from MrMsDTW, thus obtaining measure positions for all recordings.

It is possible that some recordings are going to be potentially invalid (do not follow the same sheet music), which is automatically detected and users are warned to double check this. However, this feature is experimental and does not work for all cases. Manual correction, therefore, is needed.

If we are left with valid recordings, we can play, listen to, navigate through, compare, and visualize parameters and differences. The second concept is the utilization of feature reduction method, called maximum relevance minimum redundancy. This method is applied primarily to the duration of measures and outputs first ten most relevant measures that differentiate between either two groups of recordings (which can be set by “assigning labels”) or one vs. rest scenario (measures of one recording vs. all other recordings). The software shows the ten most different measures on average and provides interface to easily select and listen to these sections in any recording. Furthermore, users can visualize differences and all computed parameters

**Sign in and Log in**

First, the user creates a free account to use the MemoVision tool. After clicking on "create an account" (Figure 3), users are guided through the registration process (Figure 4). Data are stored only in a local PostreSQL database, passwords are encrypted. No authentication is needed, so far, email could be artificial or nonexistent.

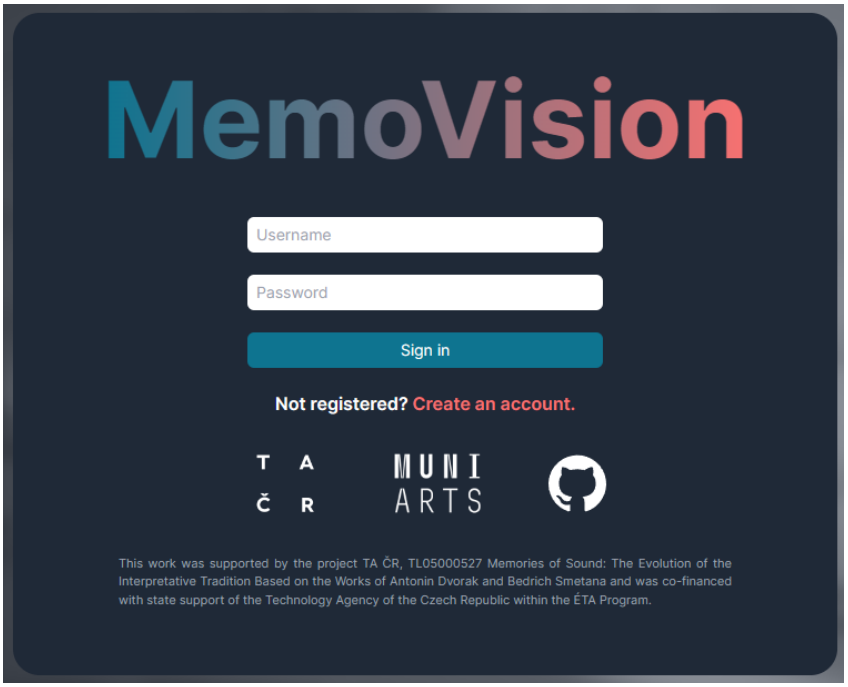


Fig. 3: The welcome page of MemoVision.


Fig. 4: The registration page.

After login, users can create sessions (only letters and numbers are accepted) to distinguish between projects and analyses (Figure 5).

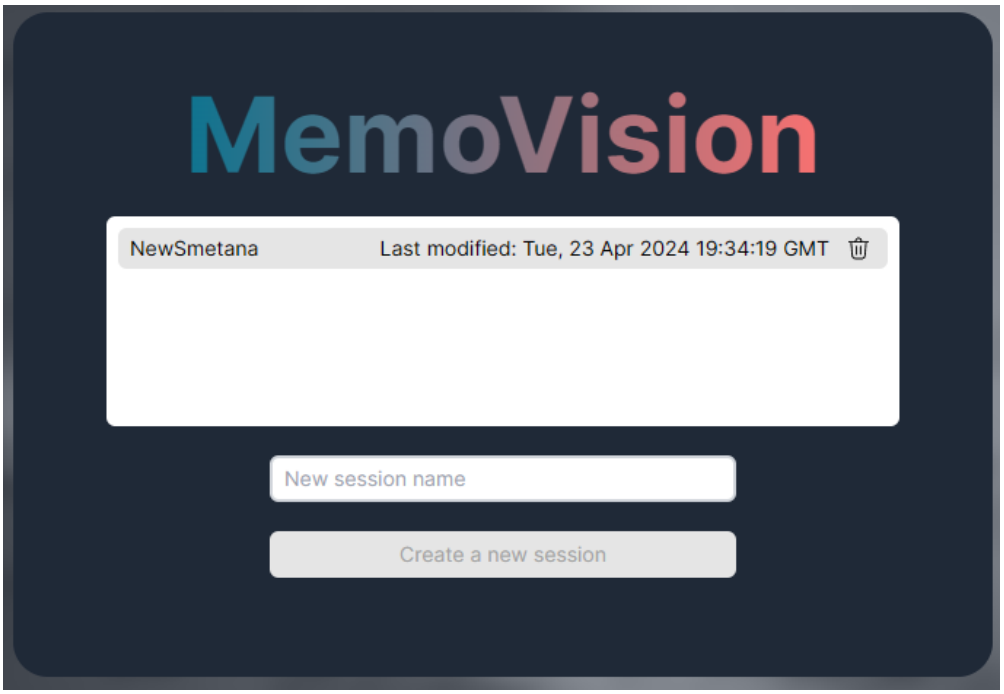


Fig. 5: The session page.

The following page is called Track manager (Figure 6). Users can upload recordings, metadata, assign labels, select reference recording, and download measures.

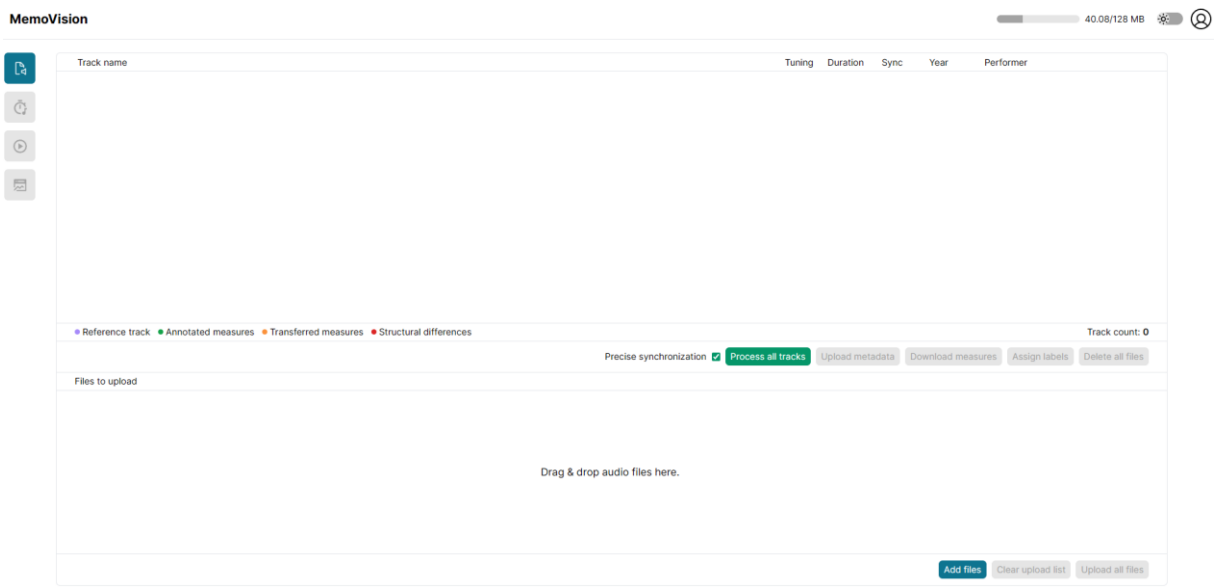


Fig. 6: Track manager interface.

On the left side, there is a panel with other functionalities – region selector, interpretation player, and visualization tabs. On the top right side, you can see your space server allocation and account information. Bottom part aims to provide interface to upload recordings and top part to manage them. After uploading several recordings and clicking on the *Upload all files*, you can add metadata manually or click on the *Upload metadata* button.

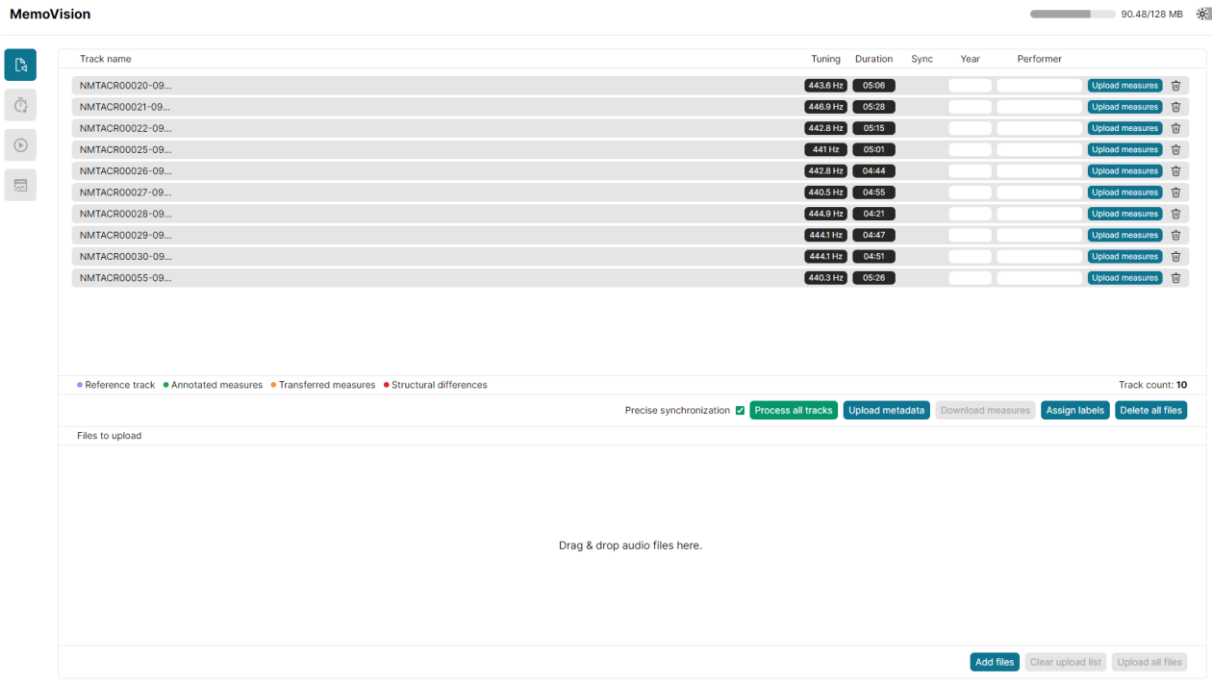


Fig. 7: Track manager after uploading audio recordings.

The metadata are not needed for the MemoVision to process data, but the *year of recording* values provide more options for comparative analysis. Next, click on one of the recordings and *Upload measures* to upload its measures in a form of .txt or .csv file. It is a simple one column (or two, see the following link) of float values corresponding to the time position of measures. For more information about the format and downbeat estimation or detection, see https://www.music-ir.org/mirex/wiki/2021:Audio_Downbeat_Estimation. To upload metadata automatically, make sure to upload a .csv or .xlsx file with a header containing filename, performer, and year columns (Figure 8). All information should be ready to process and the computation can be initialized (Figure 9). Click on *Process all tracks*. However, it can be done manually.

| | A | B | C |
|---|---|---|---|
| 1 | filename | performer | year |
| 2 | NMTACR00065-19-UC | Matějíček, Kryštof | 2023 |
| 3 | NMTACR00065-06-UC | Cetnarowski, Jakub | 2023 |
| 4 | NMTACR00065-28-UC | Šabík, Matej | 2023 |
| 5 | NMTACR00065-24-UC | Praženica, Pavol | 2023 |
| 6 | NMTACR00020-09-UC | Cechova, Jitka | 2006 |
| 7 | NMTACR00021-09-UC | Repkova, Vera | 1959 |
| 8 | NMTACR00022-09-UC | Leichner, Emil | 1994 |
| 9 | NMVISME-2324100123-02-01-UC | Maxian, Frantisek | 1942 |
| 10 | NMTACR00025-09-UC | Kubalek, Antonin | 1988 |

Fig. 8: Optional metadata in an excel .xlsx file format.

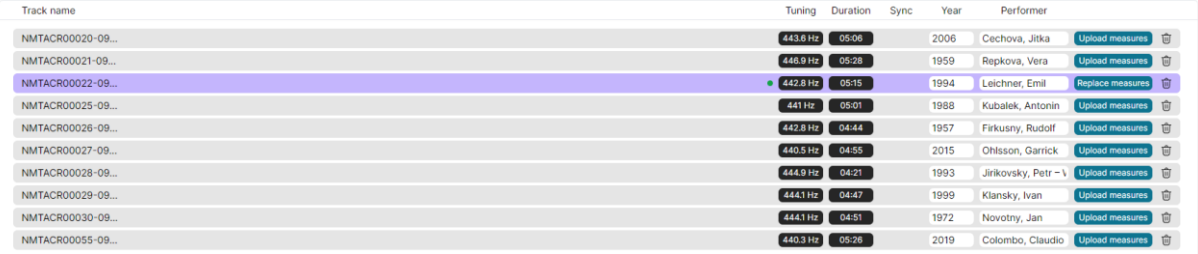


Fig. 9: All information is filled, reference is selected, and reference measures are uploaded.

The software will first find potential duplicates, computes synchronization of all tracks compared to the reference, and estimate different structures from the reference recording. For example, in our example, one track was marked as a track with one "different region" that may not follow the same harmonic structure (differ in sheet music) as the reference recording (Figure 10). We recommend keeping the tracks and personally check the validity of these tracks in the *Interpretation player* tab. Furthermore, performance parameters are computed for all given recordings (Figure 11).

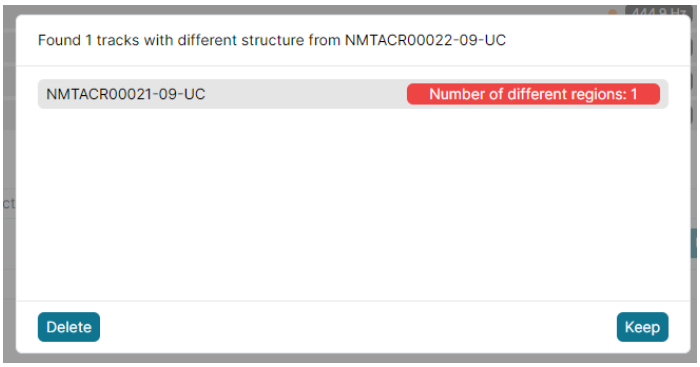


Fig. 10: Notification about possible different regions in one of the tracks.

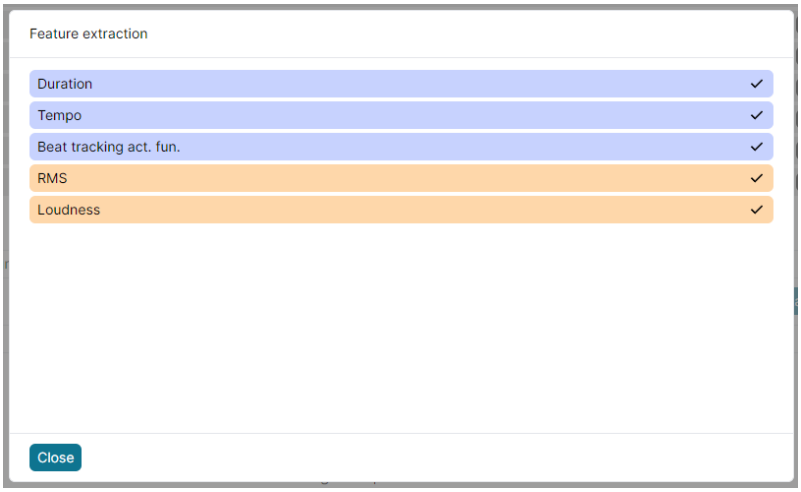


Fig. 11: Computation of performance parameters.

In the next tab, *Region selector*, you can navigate the reference recording, playback any sections and create new regions by clicking and dragging on the horizontal bar below the waveform visualization (Figure 12). If the composition is not in 4/4 timing, users can manually select portions of the piece and adjust the timing information accordingly.

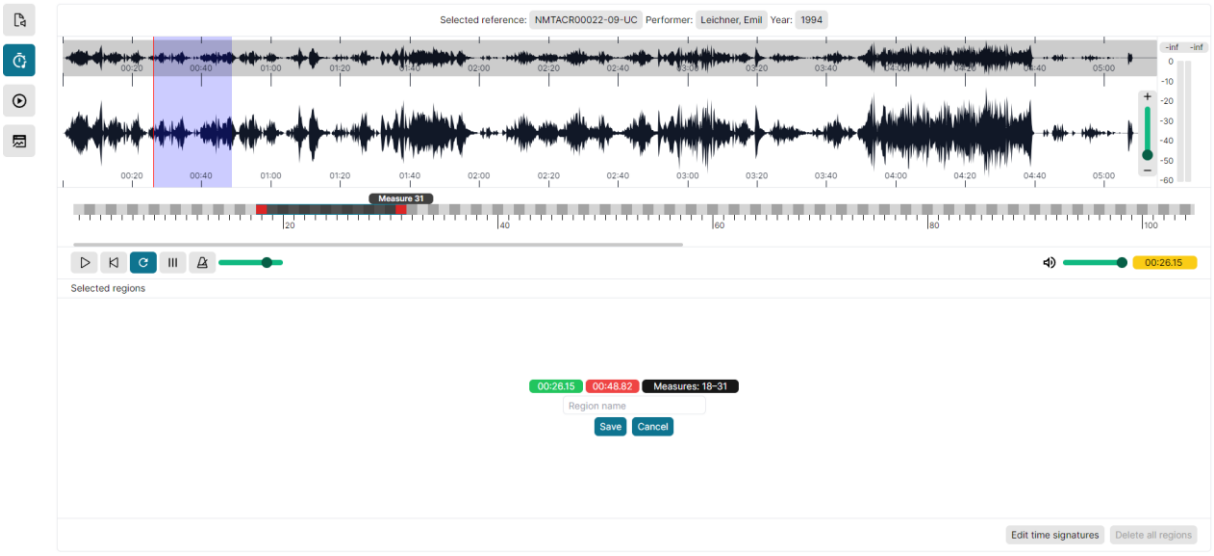


Fig. 12: Region selector interface.

The following tab, Interpretation player, contains main functionalities of MemoVision software.

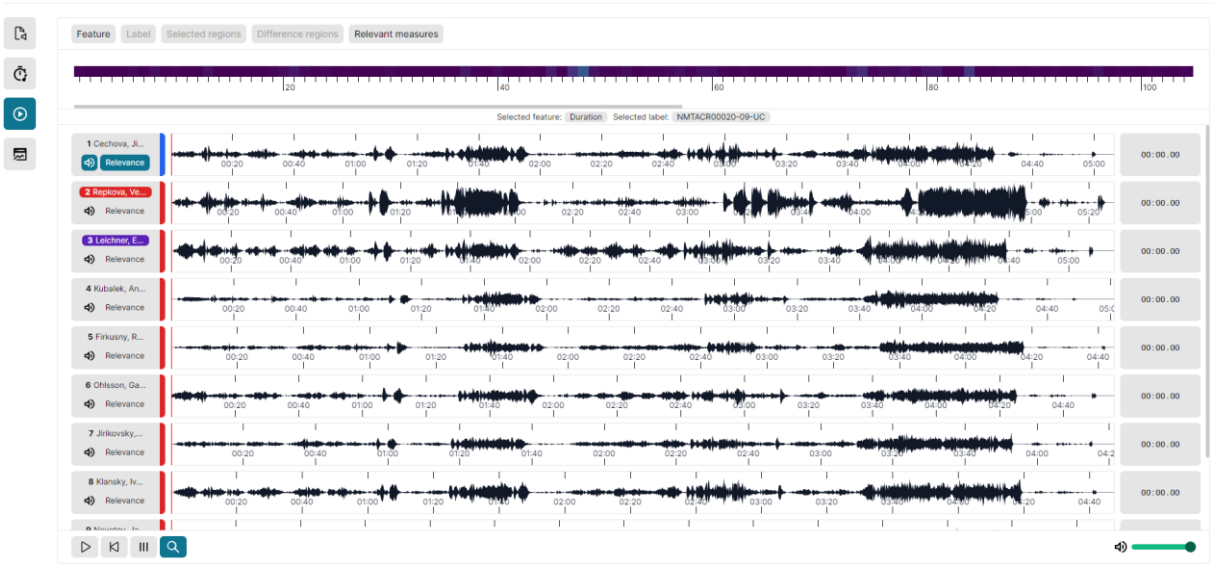


Fig. 13: Interpretation player.

Fig. 13 shows the interpretation player. In the top bar, you can select *Feature* (on which the relevance is computed, but Loudness is experimental right now), *Label* splits the recordings into two separate groups based on label information from the first tab, *Difference regions* button shows the areas where recordings supposedly do not align (has to be checked manually), and *Relevant measures* list the relevance of measures for a given composition and recordings. Then, there is a bar with the number of measures, which helps with the orientation and shows the relevance by color changes. More green or yellow means more relevance, on the other hand, darker blue means less relevance. There is a slide bar as well underneath, allowing for navigation thorough the piece. Users can select any range of measures by clicking and dragging the mouse over the bar and corresponding measures, which will select these measures in the player itself.

Next, you have the names of the recordings on the left side. If the speaker icon is highlighted, then the recording is now active (it is going to be played). If we click on the *Relevance* button, relevance is computed as "one vs. all", or in other words, this recording is compared to all other recordings based on the relevance of given features (mostly duration of measures). The color indicates two selected groups of recordings (given by one vs. all scenario or labels). In the bottom bar, there are play/pause buttons, back to the start, show measures and zoom options (we recommend using mouse wheel instead).

The last tab is the Visualization tab (Figure 14). You can select *Features*, *Labels*, *Measure features*, *Selected regions*, *Regression plot*, and *Show chords* buttons. *Features* buttons stand for visualization of individual parameters. They can be toggled and show new plot for every recording, as depicted on Fig. 14. Currently, there are RMS (dynamics to some degree), Loudness (subjective based on human hearing), and the beat tracking activation function (peaks or local maxima should correspond to the beats). *Label* button changes the left side of the recordings and show their respective colors for given labels. *Measure features* show parameters in a new plot, but its x-axis is computed in measures instead of physical time. This is crucial for analysis. Now you can compare given features not in time of one recording, but in time positions shared by all recordings (the same position in sheet music). *Selected region*s preselect the regions created in the second tab*, Regression plot* shows a new plot with regression analysis, and *Show chords* button shows chords for all recordings (experimental, may not always work properly, especially with complex music).

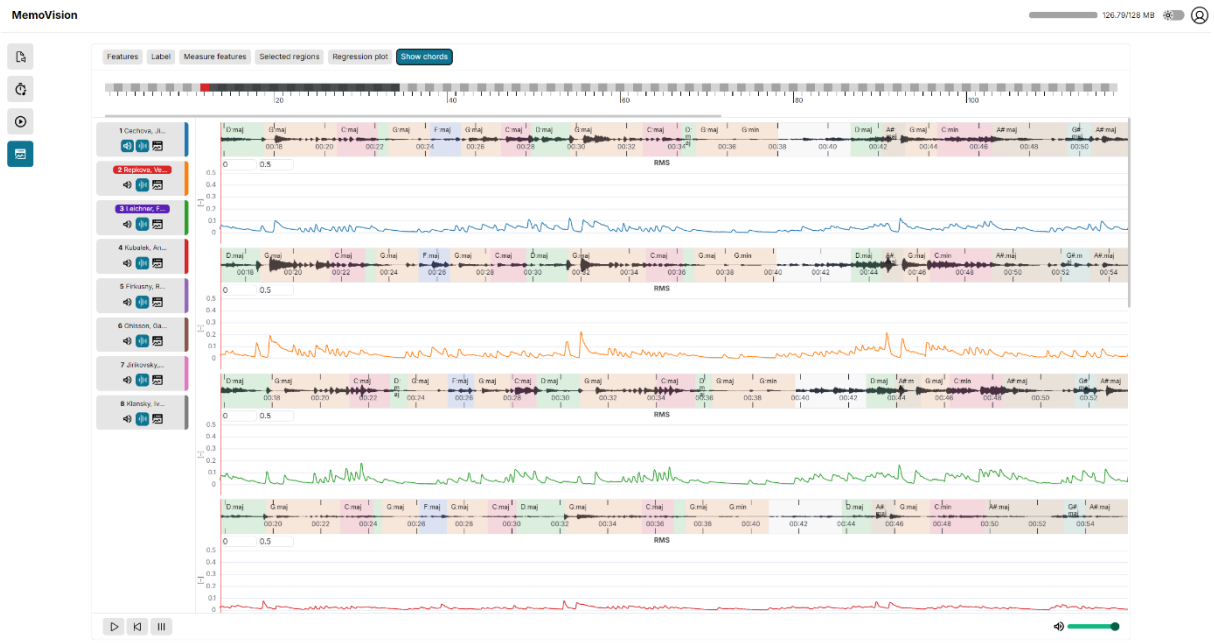


Fig. 14: Visualization tab.

The regression plot is shown in Fig. 15. On the left panel with the names of the recordings, you can select, which recording should be played, if the recording should be shown, and if the recording should be included in the regression analysis. When labels are selected, it separates all recordings to according groups and compute trend (how it evolves, statistically, in time). To be able to support claims based on this plot, one needs a lot of recordings! Note that this plot is available only if information about years of recordings was provided.

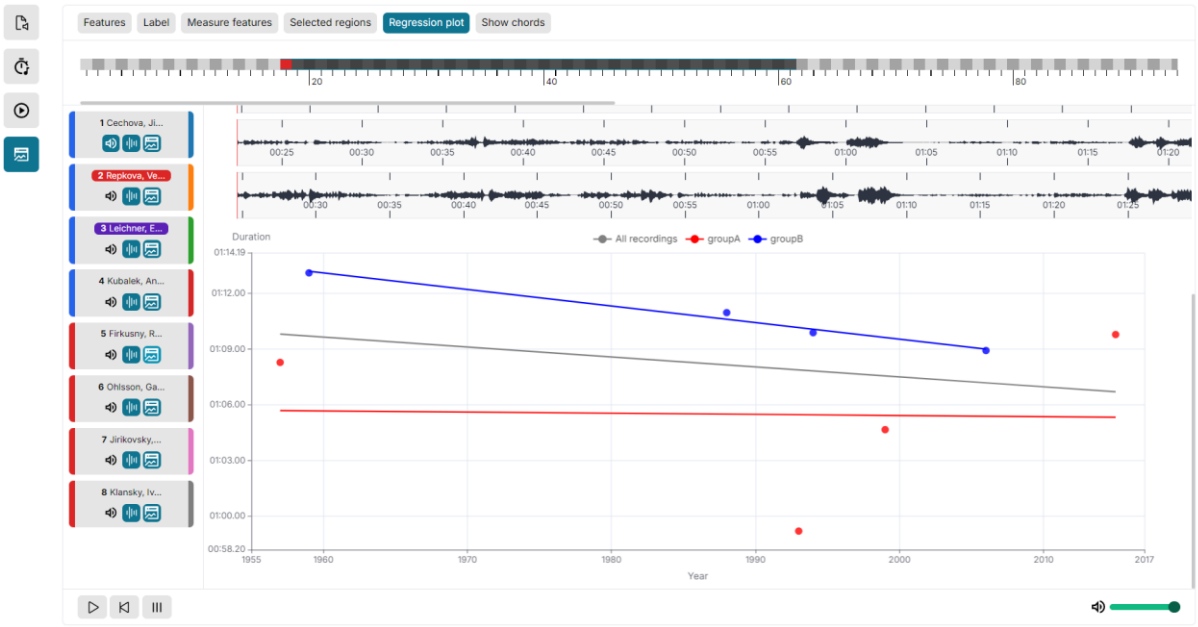


Fig. 15: Visualization tab with regression plot.

Right now, export of plots is not natively possible. Feel free to experiment with the software, run it locally from git repository, create issue on git, or ask questions at matej.istvanek@vut.cz.