Milestone 4 -Final Report

1. Motivation

The focus of this project lies on the real-time heart beat visualization of six dancers during a ten-minute ballet performance. The heart beats are measured through chest belts worn by all dancers and displayed as a typical EEG. These logged entries of the heart beats may vary throughout the piece according to the choreography as performers dance alone, form couples or dance in groups. The data gathered from this performance should be meaningfully visualized in order to enable further exploration of it. What should also be mentioned is the fact that our ideas are based on the assumption that the physical difficulty of each dancing phase is directly correlated to the intensity of the heart rate at each corresponding point.

Apart from analyzing the particularities of the various heart beats, the project will tackle technical errors that might have occurred during this performance. These issues will also be revealed by the heart beat entries recorded during the piece and were probably caused by the following factors. The close proximity of two dancers may have resulted in more than one device recording the same signal, referred to in this report as signal overlapping. Increased effort on the other hand, was the cause of signal dropouts.

1.1 Tasks

When coming up with ideas on what charts to create and how to display them, we had in mind the following three tasks as a guideline.

- How often and when is the heart beat missing?
- What heart rhythm patterns can be found?
- What are the differences and commonalities of the six people's heart rhythms?

1.2 Users

Our visualization should be suitable for any user that has a slight interest in this topic, being able to get an insight into it. The visualization is designed in such way, that it enables users to quickly gather various information regarding each dancer watching the overall progress of the heartrates. Although, when going further into this subject, those who would profit most from it are the people directly involved in the performance. Having said that, it should be mentioned that after giving it some thought, we decided upon two target user groups, whose interests will be specified in the following.

The first target user group includes people directly involved with ballet or any other dancing activities, such as choreographers. A choreographer is responsible for the outcome of the performance and for organizing the dancers in such way, that they resonate with their partners and best communicate with the audience through their movements. This is why it is important for a choreographer to first get an understanding of how difficult and tiring the performance is perceived by the dancers, in order to adapt it accordingly. Also, given the fact that the performers often dance in pairs, it is necessary that they are compatible with each other, so after analyzing their heart rates, the choreographer can pair them according to their physical capabilities and offer them parts that are more or less demanding.

The other target group is represented by individuals responsible for the technical aspect involved in the performance, namely the implementation of the heart beat measurement. As there are some errors when it comes to the transmission of the signal recorded by the dancers' chest belts, a technician might be interested in resolving this problem. In order for him to find a solution, it is important to look into any factors that might have caused the problem. Analyzing the charts, a technician will understand the severity of the problems, being then able to identify if they were related to any aspect of the performance and finally make necessary improvements to the devices, so that this kind of problem is avoided in the future.

1.3 Data

As far as the data goes, there were to datasets put to our team's disposal, one of them, namely the one containing the heart beat logs, being of more help to the creation of our final visualization.

The main dataset containing the logged heartbeats of the ballet dancers throughout the whole performance is only composed of two fields. Each line represents one heartbeat. For each heartbeat there is a timestamp given and a capital letter representing the dancer it belongs to. As these bare timestamps do not provide enough information to be visualized, they have been transformed into a more meaningful measure, which quantifies heart-beat rate (how fast the heart beats) in time. The resulting unit we went with is BPM - beats per minute.

The other dataset contains coordinates of the dancers during the performance. It is comprised of six fields, the more meaningful fields containing the X and Y coordinates (ranging from zero to 21) and a number representing the dancer, as well as the second this information was logged. The reason this dataset wasn't eventually of much use was that the files only contain dancers' coordinates for one/two minutes (the performance being ten minutes long), which is not nearly enough information to generate charts in a way that transmits the key information.

1. Related work

What we used most, was charts that we have seen in sports apps, together with the feedback we were given for our Lo-Fi prototypes. From the beginning, we knew we had to find a way to detect the dropouts of the signals which was a challenge. As people who practice sport, we see charts with average HR (bpm) every day and that helped us to have an idea of what the graphics should look like. Runtastic and Health (application preinstalled on Apple devices) are two applications from which we've inspired. To detect the dropouts of the signal we have used similar graphics but with broken lines. The tree screenshots below (Runtastic and Health) represent the graphics from which we started. Although the graphics from these apps focus more on the general overview of the data as a starting point was perfect. With this concept and after a brainstorming season we have come to create 10 different charts, that includes an animation, a heat map, a box-plot and many line charts. The feedback we were given was especially important to find the dropouts of the signal. As academic and commercial tools we have used unet hosting, Php, Excel and Tableau.

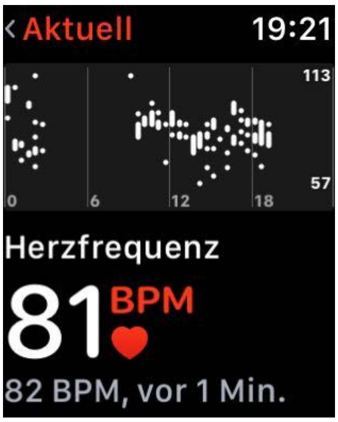

Figure 1

Figure 2

Figure 3

2. Approach

We have decided to approach a design study project using Tableau. However, to extract and calculate the BPM we have used more two tools: Php and Excel.

Our storyboard consist of 3 dashboards: General overview of the data, Error analysis and Spatial analysis.

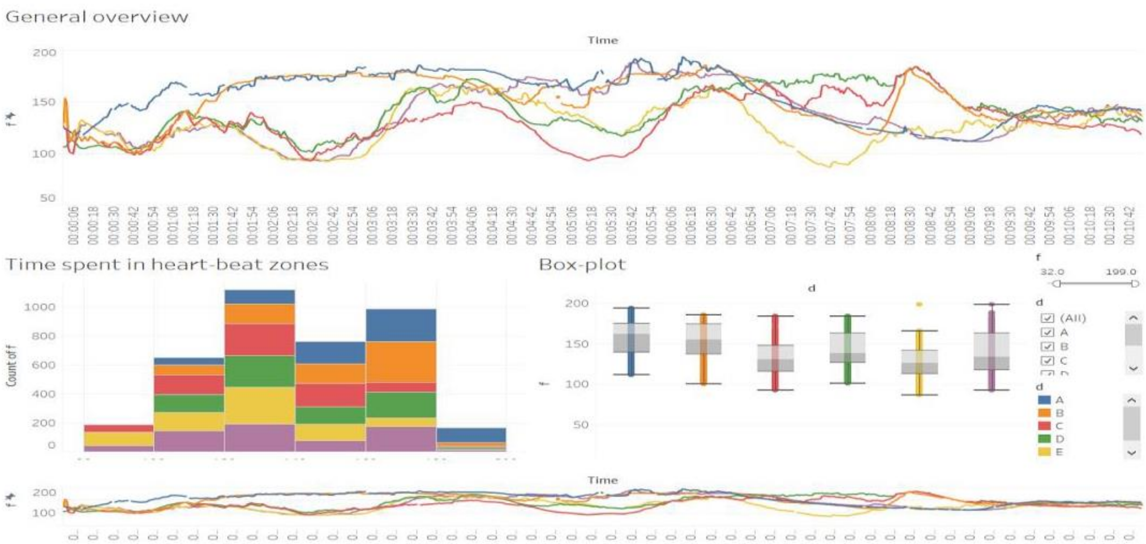

Figure 4: General overview of the data

The aim of this dashboard (figure 4) is to provide the user with a good idea, of how the data looks like, how it is distributed overall and indicate if there are any trends and correlations present.

At the bottom of this dashboard, the user can use the brushing tool to select certain data-points. Every data point can be of course hovered, to see the corresponding tooltip - this is valid for all our dashboards. As can be seen in the dashboard, we use a line chart, a box-and-whisker plot and a bar chart. We chose these types of charts in order to better interpret the data. As mentioned in 'Related work', sports apps have been some of our inspiration sources and the line chart is one of the results. In this chart, the heart-beat rates of all the dancers are visualized, distinguishing them by color. The x-axis represents the time. The user can quickly observe possible correlations between the dancers. Sometimes, there are missing signals in the data, resulting in a frequency drop. The visualization of the drop is omitted in this chart, because it is not relevant for the trends and it would only overcrowd the chart. Box-plots (top is a maximum value; bottom minimum; box in the middle represents 50% of the data; the color-border in the box is median) help us a lot in understanding data, we used box-plots in the statistic courses and this was one of the reasons for choosing this chart. The bar chart is the result of the brainstorming session and the feedback received at the M2 presentation (Lo-Fi Prototyping). On this chart, we can see the overall distribution of the data for each dancer.

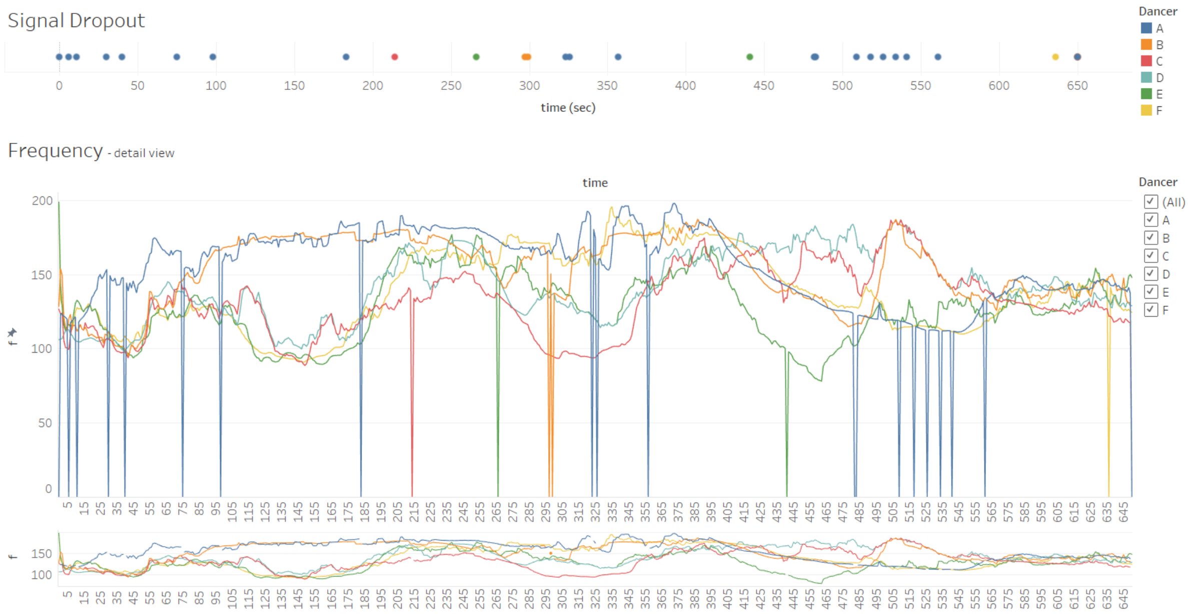

Figure 5: Error analysis

The main chart in this dashboard is a line chart, displaying the heart frequency for each dancer over the time. Although this chart may seem slightly crowded, showing the lines where the frequency drops is of great significance in this case. As this dashboard's main focus lies on the technical problems, this is rather a necessary step, than one that confuses the user. The chart displaying the signal dropouts supports the main chart, enabling the user to quickly see when and whose signal got lost, respectively how often did each signal drop.

The dashboard makes use of some visualization techniques, such as linking, brushing, filtering, and tooltips. The chart at the bottom of the dashboard represents the context view for the frequency chart. Being linked to the other two charts, it enables the user to brush over an area (in the context view) that he is more interested in and see changes in both the frequency detail view and the signal dropout chart. Whenever hovering over any point in a chart, there is a tooltip displaying relevant information for that point. The reason for choosing these line charts was mostly the feedback that we have received during the implementation.

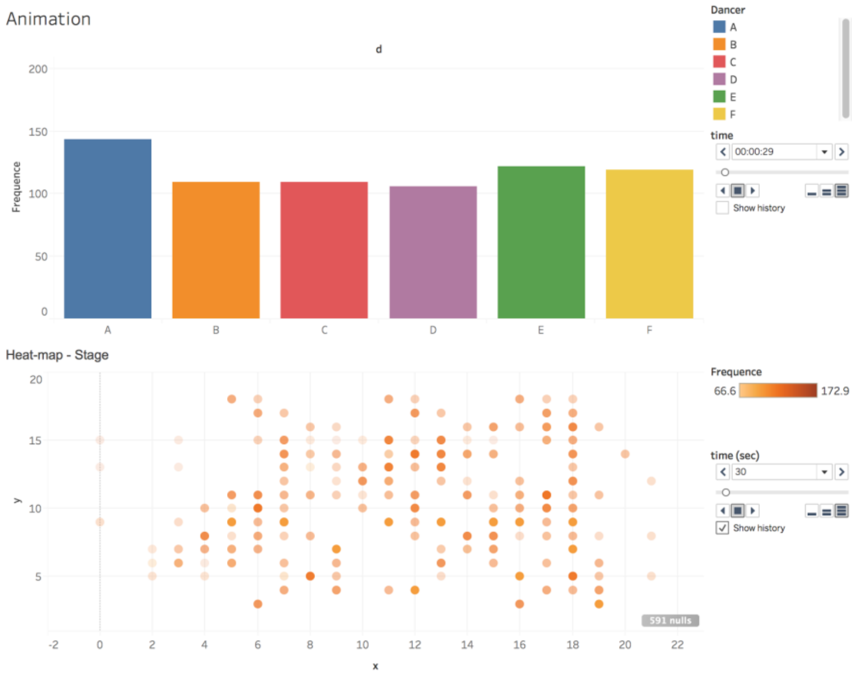

Figure 6: Spatial analysis

This dashboard contains two charts: an animated bar chart and a heat-map. The main reason for the choice came from the feedback review. We initially thought about making an animation with circles. The higher the pulse, the larger the circle's size. However, we eventually realized that such animation would have created a lot of confusion, so the circles have been replaced with bars. Each dancer is represented by a bar and each bar changes its size according to the BPM. The higher the pulse, the higher the bar. In Tableau, we found a very good widget for our animation, which enables the user to select the speed of the animation (how fast the time passes). For example, the user can see the animation for 10 minutes and 50 seconds (standard), or he can use the button "fast" and the almost 11 minutes of dance will be transformed in 1 minute and 5 seconds of visualization. In this way, the user can see the increase or decrease of the heart rate more easily.

The Heat-map represents the stage where the ballet dancers dance for 10 minutes. Each point represents a heartbeat in that location. If the opacity of the points is high, it means that the heart rate is higher, in our case the orange color is more prominent. The reason for choosing this type of chart is obvious, given the fact that it uses x and y coordinates.

3. Implementation

As mentioned in the approach, to produce the visualizations we have used Tableau. As challenge was not only to create the charts, but the data from the dataset had to be calculated and converted into a compatible format for Tableau, like CSV. To do this, we have used PHP and Excel. Another challenge was to create a webpage. Our team decided to create the webpage using HTML and CSS, two technologies we are familiar with. We have decided to use Tableau and not other platforms such as Siense, Tibco Spotfire or D3 due to multiple features it offers. With Tableau, it was even possible to create an animation that shows the heart rate frequency of the ballet dancers as separate bars, whose height increases and decreases for 10 minutes according to their heart rates.

4.1 Challenges

Data preprocessing was one of the first challenges we encountered. To preprocess the dataset we used PHP. The data comes in a .csv file, with the features t (time - the whole number representing the second elapsed), f (frequency), d (dancers label) and T (period). The computation has been done in the following way: we first determined the periods in between the heart-beats, which we then inverted into frequencies, and for each second of the performance we computed one frequency record. Additionally, for the frequency computation, we used a moving-average filter, in order to obtain as reasonable values as possible. This means, that the frequency is computed based on the last X (= window-size) heart-beats. The best window-sizes for the filter lie approximately between 30 and 50 records, we think. To determine the BPM we use the following formulas:

$\overline{BPM}_{n,w} = \frac{1}{w}\sum_{i=n-w}^{n-1} \boldsymbol{BPM_n}$ *; n = sequence index; w = window size*

$T_n = t_n - t_{n-1}\ ;\ n \in N, \forall n > 0$

$f_n = \frac{1}{T_n}[s^{-1} = Hz] => BPM_n = \frac{1}{T * \frac{1}{60}} = \frac{60}{T_n}[min^{-1}]$

Other challenges and problems that we encountered are mainly bound to the files containing the coordinates. The files only contain dancers' coordinates for one minute (the performance being ten minutes long), which is not nearly enough information to generate the charts we have wanted in a way that transmits the key information. For this reason, the error analysis dashboard and the spatial analysis one are lacking one chart, the one for the error analysis being a quite important one.

5. Results

5.1 Scenarios

Note: for the usage of our tool, user-attributes like name, sex or age are not relevant. Therefore they are not specified.

Target group 1:
Ballet /dancing related people

Choreographer, who wants to obtain an overview on the physical difficulty of the performance. For this persona, the following questions are a subject of interest:

- How much do various dancers utilize their physical capabilities comparing to one-another
- How well are various roles in the performance (un)balanced
- Is there a need to reorganize the pairs according to their physical capabilities (based on the assumption, that the heart rate directly correlates with the physical condition of the dancer)

Getting our tool involved:

The first thing the user sees, is an overview of all the present frequency-data:

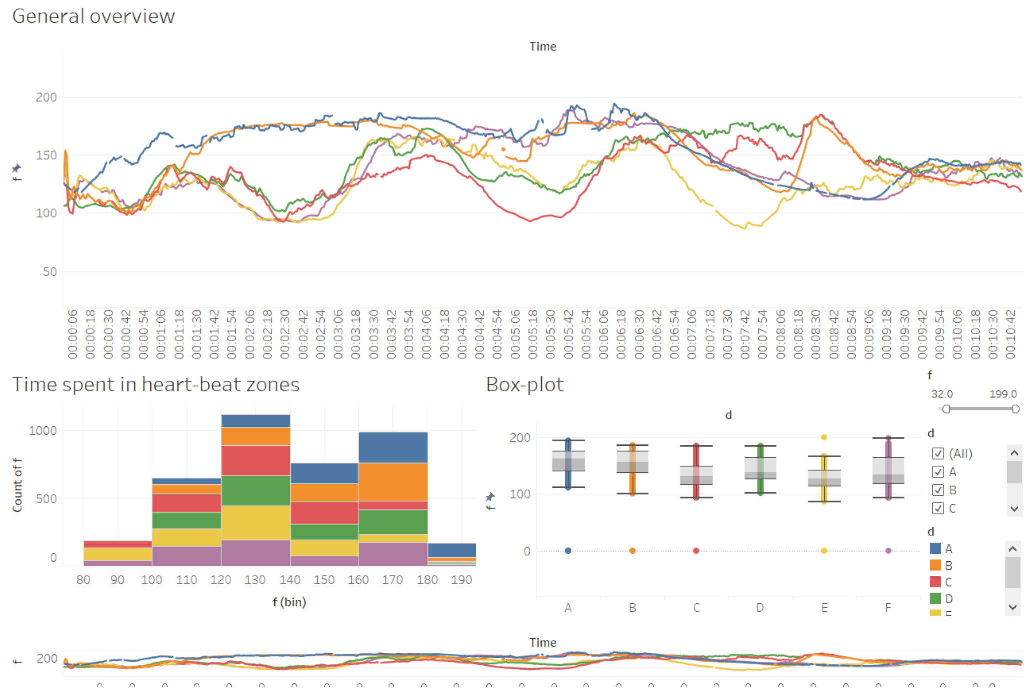

Figure 7: On the first glance

The user starts analyzing the first chart. Two major trends are visible – these indicate a possible similarity in the choreography of the involved dancers. One "cluster" would be dancers A and B (blue and orange), and then the other dancers (C-F):

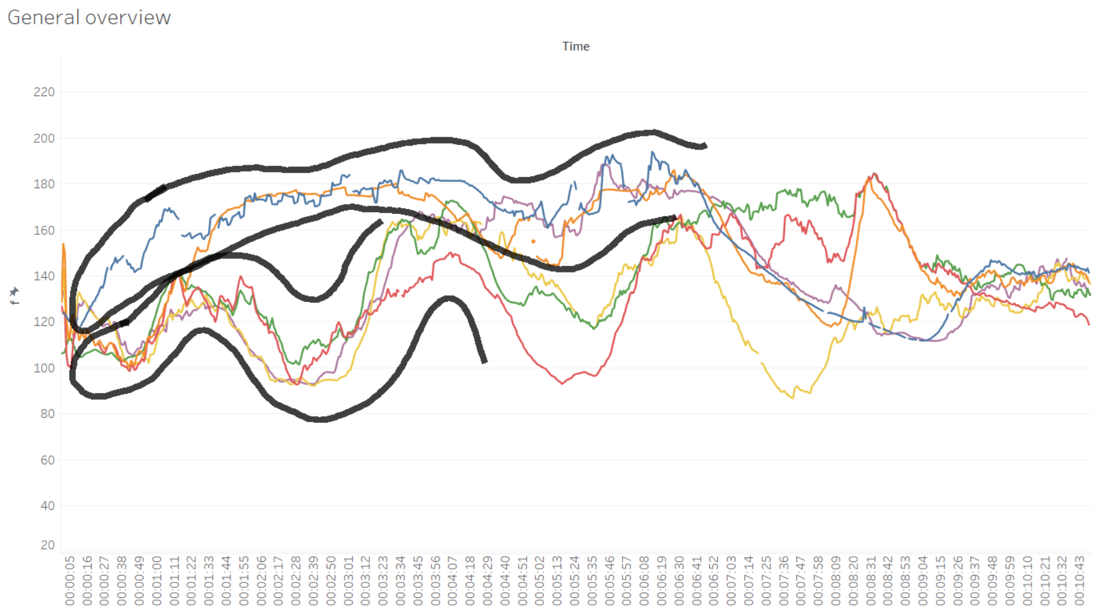

Figure 8: looking for correlations, similarities, clusters

The user then isolates the dancers A and B from the other users, to see their individual comparison. This is done with the filtering widget:

Figure 9: Filtering options

The user then obtains the following view:

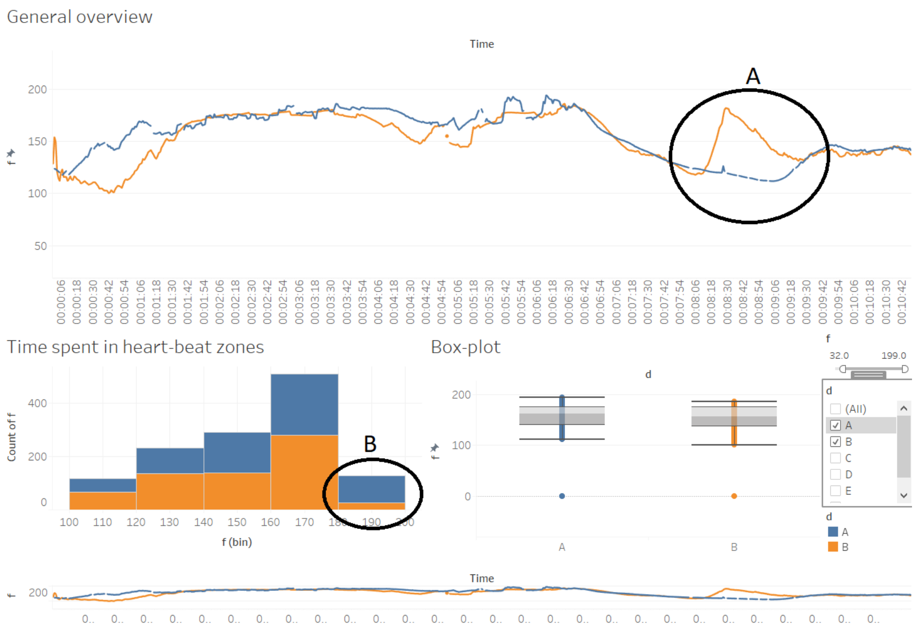

Figure 10: A and B isolated

At the first sight, the utilization of these 2 dancers seems to be similar. If our assumption is true and these dancers share common features of the choreography, this then indicates a well-balanced pair.

However, the choreographers bright sight notices, that overall, the *blue* dancer's heart rate tends to be higher, than the one of the *orange* dancer. Especially, in the highest heart-rate zone: 180-200 BPM, the blue dancer spends significantly more time, than the orange one (annotation B). This can be caused by different physical condition of the two dancers, but also by many other factors. The choreographer may further investigate attributes like sex and age (of these dancers), which also contribute to the heart-rate differences. A downside to the tool is, that this kind of personal data is not present, and therefore it was not considered for incorporating and visualizing.

In the end, the choreographer may state, that the roles of A and B are pretty well balanced and there is no big need for rearranging the choreography because of these two dancers. The before-mentioned aspects (sex, age) may be inspected by the choreographer, in case of interest.

The second anomaly, which the choreographer notices, is the sudden jump (annotation A) of the orange dancer's heart-rate. This indicates a signal-transmission problem, because such a quick heart-rate growth is not very trustworthy. He may and may not further investigate this problem, in the way the **technician** does it later. Also, the choreographer may analyze the other group of dancers in the way, he/she did it for dancers A and B.

Target group 2:

People responsible for technical implementation of the heart-beat measurement

Technician, who wants to optimize the heart-rate signal transmission to the best possible level. A possible subject of inspection may also be trying out different hardware devices for this purpose, and observing their performance. For this persona, the following questions are a subject of interest:

- How often and when does the signal drop out?
- How often and when does a signal-overlapping (receiving the same signal by multiple transmitters - dancers) occur?
- Who (which dancer) tends to have a lot of signal transmission problems?
- Are there more problems at higher heart-rates?

Getting our tool involved:

Opens the tool and sees the same view, as it is shown in the first figure of the previous scenario. Switches to the “Error analysis” view, and sees the following:

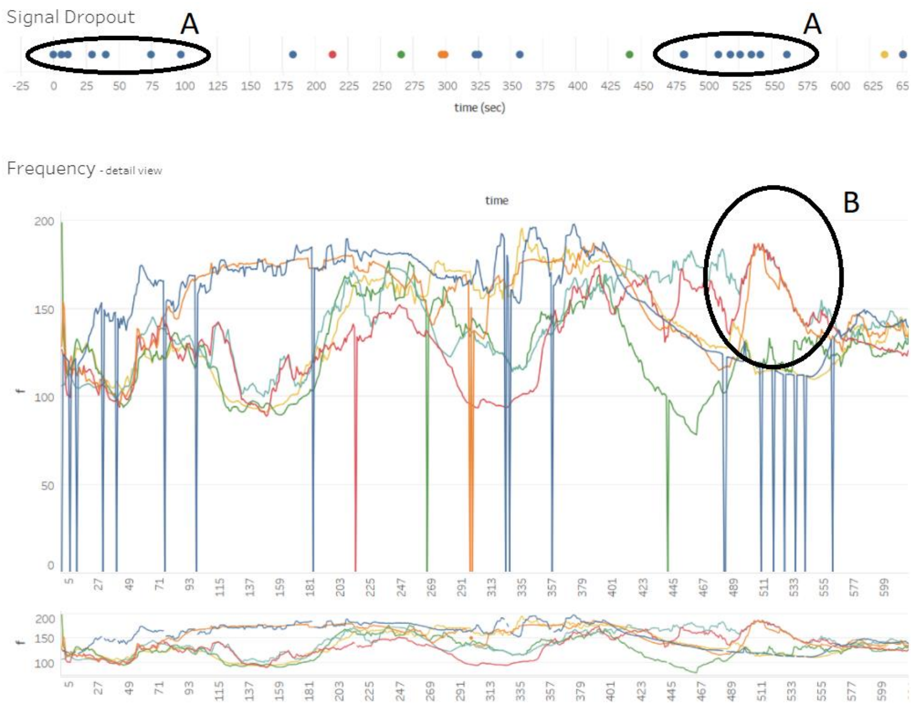

Figure 11

At the first glance, the technician notices, that the blue dancer has significantly more signal dropouts, than the other dancers. Moreover, the dropouts are suspiciously located at the very beginning, and near the end of the performance (annotation A). Dropouts of the other dancers are not so critical.

After a closer look on the middle view, the technician notices a suspicious equality of BPM rates of multiple dancers (annotation B). This indicates the signal-overlapping problem (receiving the same signal by multiple transmitters). The technician then focuses on the segment more precisely:

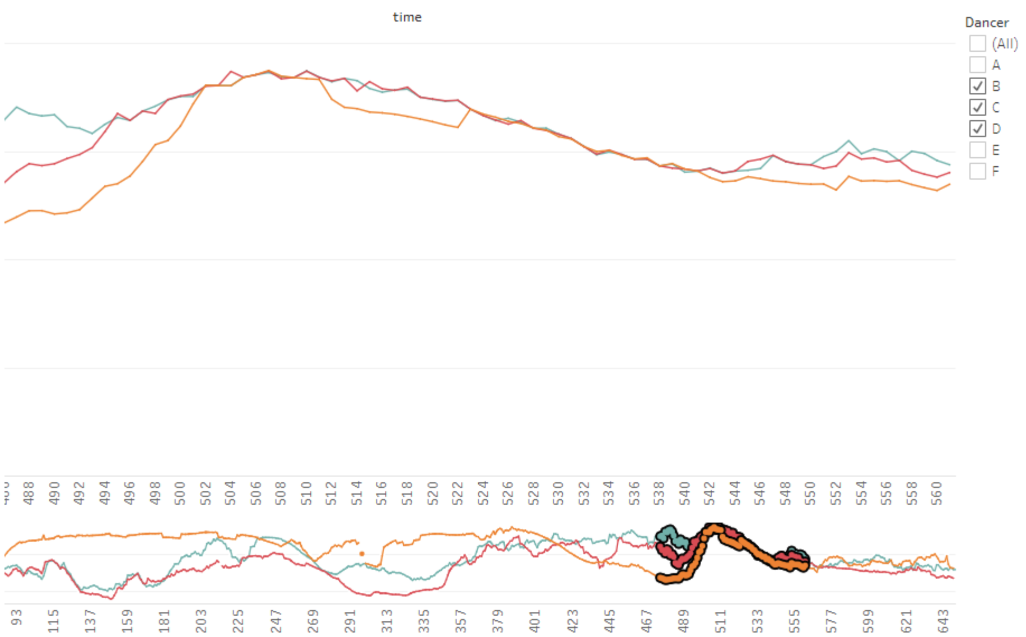

Figure 12

He/She does this by brushing over the desired time-period in the bottom view, and then isolating the dancers B, C and D, which all happen to receive the same signal. This is obvious to the technician, because the exact same trace on an approximately 40 second long interval cannot be an accident. The exact same line is in fact probably even a little bit longer, than shown on the picture - the moving average filter causes a lag of couple of seconds (the actual signal overlapping started probably ca. 10 seconds before it is visible in the visualization).

If the tool contained reasonable location data, the technician would normally right now switch to the 3rd view of the tool, and inspected, whether the dancers were close to each other at the time period, where a signal overlap is suspected. However, the location data, we worked with, is very limited and in this case does not provide additional information for this use-case.

The technician obtained an idea, of:

- When did a signal overlap happen,
- Who did it happen to

He/She can now further investigate the circumstances on the stage during the performance, and maybe get a better clue for diagnosing the fault (why does the blue dancer have so many dropouts?).

5.2 Performance of the system

Our visualization tool is done with tableau, and therefore can be used in two ways:

- Embedded into a website (http://homepage.univie.ac.at/hronskyr97/VIS/M1/m3.html)
- Opened with Tableau

When working with the browser, the performance is slightly worse than in Tableau. However it is possible to use it that way, and if doing so, we encourage you to work in full-screen mode, for a better experience.

When it comes to speed, the tool performs well, mainly because our dataset is not too big.

We designed our tool while keeping in mind, to logically separate the parts of the tool as good as possible. We ended up with a concept of 3 dashboards, for:

1. Gaining quick first impression, getting an overview of the overall distribution of the data
2. Focusing on error analysis of the data
3. Focusing on spatial analysis

In each of the dashboards, we tried to always give the user the ability to look at the data from a "higher" and a "lower" perspective. For the higher perspective is mainly responsible the main frequency-in-time chart, which can sometimes also be brushed over to select a certain time-interval only. With the other charts, we always tried to emphasize the very purpose of the dashboard (→ data-distribution, errors, proximity).

Also, every dashboard contains filters, for focusing on particular dimensions of the data. The best example for this is filtering the dancers.

With the tool assembled this way, the user should be able to quickly achieve the goal, without losing orientation and eventually being frustrated.

5.3 Feedback

In general, we received a rather positive feedback. Based on a very early feedback, we avoided using a pie-chart straight from the beginning (M2), and switched to a stacked bar-chart (for heart-rate zones). After M3, we were advocated to use color-blind friendly palette for categorical data (dancers in this case). The original version had the default tableau colors, so we did not think about something being wrong with them.

We also received the following feedback:

> *"I wonder if you could do a line chart that would gradually reveal as the performance went on for the heart beats though. The advantage is that you'd have some history of the heart beats."*

We decided not to do it this way, because the historical context is already present in the main *frequency-in-time* chart, and there is not really a need to animate it. The user can still brush over it and focus on certain time intervals. The "bouncing bar-chart" however nicely emphasizes the current ratio/difference between the BPM-height of different dancers (which is not so clearly visible in a line-chart). That is the intended purpose of the animated bar-chart. In a combination with an animation, this gives the user a nice way to go through the whole dataset – manually choosing between approximately 600 frequency data-points would be tedious.

6. Discussion

6.1 Strengths and weaknesses

Strengths:

One of the advantages of our solution is, that the user gets an overview of the whole performance very quickly. In fact, the main frequency-in-time line-chart may be also interpreted, as some kind of "physical difficulty profile" of the performance. The user can easily determine the physical utilization of various dancers, in different parts (time intervals) of the performance.

Also, it is very easy and straight-forward to spot the transmission errors, and identify, when they occur.

As another advantage of our implementation, we consider its very easy maintenance and deployment – which is thanks to using Tableau.

Weaknesses:

Our whole approach is based on an assumption on an individual: *a direct correlation between the heart-rate and physical utilization during a ballet performance exists*. This is true to a large extent, however, there are also lots of other factors (possibly age, sex, etc.), which may create differences. It might be interesting, to have these factors implemented and visualized as well. However, we did not have this sort of data – therefore, it is not possible to observe their influence on the performance.

With our tool, it is not really possible to accurately diagnose problems – the user only finds out whether there are some problems, and might get pointed into a direction, of where, when and why the problem occurs.

In the current implementation, as already mentioned, the location-data is very limited. Our original intention was, to somehow use this data for verification of various assumptions, like:

- Signal overlapping happens due to dancers currently dancing close to each other
- Dancers with similar heart-rate profile have similar features in their choreography

However, we did not manage to fulfil this task to the desired extent, due to the lack of data. The third dashboard of the tool, which was mainly intended to deal with this problem, might therefore appear a little bit confusing to the user (with the current dataset).

6.2 Lessons learned

One of the lessons we learned is that it might be better to start prototyping, only when you already have the whole dataset, in order to avoid planning tasks which cannot be completely fulfilled. This means it is important to remain flexible and always be prepared to face any challenges that might arise. Adaptability also refers to the ability to make slight changes when necessary, such as after receiving feedback from supervisors or possible users. Lastly, what was yet again confirmed to us is that communication is the key, all of us having worked together to give life to our ideas and come up with a final visualization we all are satisfied with.

7. Task Separation

Catalin Vasilescu: Related work, approach, implementation

Rastislav Hronsky: Results, discussion

Sonia-Ruxandra Suciu: Motivation, approach

8. References

https://runkeeper.com - Runkeeper, sports-tracking app
https://www.apple.com/ios/health/ - Health iOS app
http://people.duke.edu/~rnau/411avg.htm - Moving Average Filtering
http://php.net/ - used for data processing

http://homepage.univie.ac.at/hronskyr97/VIS/data/ - PHP source code for data processing
http://mkweb.bcgsc.ca/colorblind/ - color palettes for color blindness