KAVAGait: Knowledge-Assisted Visual Analytics for Clinical Gait Analysis

Journal:	*Transactions on Visualization and Computer Graphics*
Manuscript ID	TVCG-2017-02-0028.R1
Manuscript Type:	Regular
Keywords:	Design study, Interface design, Knowledge generation, Knowledge-assisted, Visualization, Visual analytics, Gait analysis

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.
FinalVideo.mp4

KAVAGait: Knowledge-Assisted Visual Analytics for Clinical Gait Analysis

Markus Wagner, Djordje Slijepcevic, Brian Horsak,
Alexander Rind, Matthias Zeppelzauer, and Wolfgang Aigner

Abstract—In 2014, more than 10 million people in the US were affected by an ambulatory disability. Thus, gait rehabilitation is a crucial part of health care systems. The quantification of human locomotion enables clinicians to describe and analyze a patient's gait performance in detail and allows them to base clinical decisions on objective data. These assessments generate a vast amount of complex data which need to be interpreted in a short time period. We conducted a design study in cooperation with gait analysis experts to develop a novel **K**nowledge-**A**ssisted **V**isual **A**nalytics solution for clinical **Gait** analysis (KAVAGait). KAVAGait allows the clinician to store and inspect complex data derived during clinical gait analysis. The system incorporates innovative and interactive visual interface concepts, which were developed based on the needs of clinicians. Additionally, an explicit knowledge store (EKS) allows externalization and storage of implicit knowledge from clinicians. It makes this information available for others, supporting the process of data inspection and clinical decision making. We validated our system by conducting expert reviews, a user study, and a case study. Results suggest that KAVAGait is able to support a clinician during clinical practice by visualizing complex gait data and providing knowledge of other clinicians.

Index Terms—Design study, interface design, knowledge generation, knowledge-assisted, visualization, visual analytics, gait analysis.

1 Introduction

According to the 2014 United States Disability Status Report [1], 5.5% of working age adults (ages 21 to 64, amounting to more than 10 million nationwide) suffer from an ambulatory disability. Walking and stair-climbing are essential motor functions that are prerequisites for participation in activities of daily living. Disruptions to these motor skills hold severe health and socio-economic implications if left unattended. Therefore, gait rehabilitation is a crucial issue for clinicians.

Gait analysis tools allow clinicians to describe and analyze a patient's gait performance to make objective, data based decisions. The systems commonly used for capturing gait data range from simple video cameras and force-distribution sensing walkways to highly sophisticated motion capture systems [2], [3]. The latter is often referred to as the gold standard in clinical gait analysis, as this method assesses the gait pattern's underlying kinematic and kinetic components [4]. However, the motion capture system's widespread use is limited due to its substantial monetary and infrastructural costs, prolonged time commitment for data collection, and its requirement for specialized technicians. Thus, clinics with a large daily influx of patients must rely on more practical and affordable methods. Force plates and cost-effective two-dimensional gait analysis tools are popular alternatives to determine external forces applied to the ground (ground reaction force, GRF) during gait [5, pp. 83–96] and the associated kinematic variables (e.g., 2D joint angles). A typical clinical gait analysis scenario involves the following: in a first step, a clinician conducts a physical examination of the patient. Then, the patient is instructed to walk across a walkway in the gait laboratory several times, while the clinician records the patient's GRF. These analysis methods generate a vast amount of multivariate, time-oriented data, which need to be interpreted by the clinician in a short period of time. However, support for decision making based on analyzing this data is very limited in currently used systems. The resulting data are typically represented in a very simplistic manner using non-interactive visual representations such as line plots and simple spreadsheets to inform clinical decision making.

In the above-described scenario, it is a difficult task to interpret the obtained data as several parameters are interlinked and data interpretation requires considerable domain expertise. The combination of a vast amount of inter-linked clinical data derived from clinical examinations, the need for sophisticated data analysis methods, and clinical decision making requiring the judgment and expertise of clinicians, strongly lends itself to the notion of visual analytics (VA) [6], [7]. VA may support the clinician with powerful interactive visualization techniques that are integrated in combination with semi-automated data analysis methods. Consequently, this may support the clinician in interpreting complex data and drawing appropriate clinical conclusions. The clinicians' 'implicit knowledge' from prior experience is essential in the analysis process. Thus, it makes sense to externalize some of the domain experts' 'implicit knowledge' and make it available as 'explicit knowledge' in the VA process [8], [9]. As such, it can be used to augment the visual display of data and to support (semi-) automated data analysis

- *Markus Wagner, Djordje Slijepcevic, Alexander Rind, Matthias Zeppelzauer, and Wolfgang Aigner are with St. Pölten University of Applied Sciences, Austria and TU Wien, Austria. E-mail: {markus.wagner, djordje.slijepcevic, alexander.rind, matthias.zeppelzauer, wolfgang.aigner}@fhstp.ac.at.*
- *Brian Horsak is with St. Pölten University of Applied Sciences, Austria. E-mail: brian.horsak@fhstp.ac.at.*

Manuscript received M_Word DD, YYYY; revised M_Word DD, YYYY.

(knowledge-assisted VA methods). Additionally, joint learning between clinicians would be enabled as well as the collection of expert knowledge across several clinicians allows the constructing of a comprehensive clinical knowledge database [8] (refereed to as 'explicit knowledge store' (EKS) later in this article for the proposed prototype).

1.1 Overview and Method

This work follows the paradigm of problem-oriented research [10], i.e., working with real users (clinicians), and aims at solving the aforementioned problem by means of VA. In detail, a comprehensive prototype was developed which is intended to support the clinician in interpreting gait data during everyday clinical practice. The methods proposed in this work aim at externalizing implicit knowledge of clinicians into a knowledge database that makes these data available as explicit knowledge to other clinicians. For this purpose, we conducted a design study, following Sedlmair et al. [10]. Specifically, we followed the 'nested model for visualization design and validation' [11]. This model is a unified approach, which structures visualization design into four levels and combines them with appropriate validation methods, which reduces threats to validity at each level. Overall, our work contributes to visualization research in all three categories (1–3) outlined for design studies in [10] as well as presents new knowledge-assisted visualization approaches (4):

1) **Problem characterization and abstraction:** A common language and understanding between domain experts and VA researchers was established. Specific data, user, and task requirements for clinical gait analysis were set as prerequisites during the design process (see Section 3) for the development of a novel knowledge-assisted VA system.
2) **Validated design:** The design rationale and implementation details (see Section 4) were validated by conducting expert reviews, user studies, and a case study (see Section 5). The primary aim was to obtain information whether or not the developed system is of valuable contribution to the domain.
3) **Reflection:** Insights gained during the validation process (see Section 6) were reflected and analyzed to propose directions for possible future development.
4) New **knowledge-assisted visualization approaches** were used to generate easily understandable 'Graphical Summaries' of the data as well as the novel 'Interactive Twin Box Plots' (ITBP).

2 Related Work

From a data perspective, gait measurements are multivariate time series. To visualize and analyze such data, a variety of different visual analytics (VA) approaches have been introduced in earlier work.

Visual Analytics for Movement Time Series: Andrienko et al. [12] give a broad overview how VA can be used to visualize locomotion, which they refer to as 'Visual Analytics of Movement'. In their work, they give recommendations on how such data can be represented in the context of VA and how these data may be resampled. However, they mostly focus on geospatial datasets in relation to time. In the field of sport science, two VA systems [13], [14] support soccer analysts in analyzing position-based soccer data at various levels of detail. Janetzko et al. [14] additionally enrich the analysis with manually annotated events such as fouls and suggest further candidate events based on classification. An effective fully automated method for human motion sequence segmentation for character animation purposes was introduced by Vögele et al. [15]. They described the fast detection of repetitions in discovered activity segments as a decisive problem of motion processing pipelines. For testing this method, they used different motion capture databases and visualized the results with stacked bar charts for comparison with other techniques. In the context of medicine, sports and animation, the 'MotionExplorer' system [16] enables the exploration of large motion capture data collections represented as multivariate time series. Following an iterative design approach, Bernard et al. [16] demonstrated the functionality of the 'MotionExplorer' through case studies with five domain experts. A similar approach, the 'MotionFlow' system [17], allows more specific grouping and analysis of patterns in motion sequences. Another system was described by Purwantiningsih et al. [18]. They collected data on patients' quality of movement using serious games and different motion sensing devices. To make these multivariate time-series data accessible to clinicians, their VA solution allows hierarchical clustering and navigation in time. The VA system 'FuryExplorer' [19] improves analytical workflows for evaluation of horse motion by interactive exploration of captured multivariate time-oriented data.

Visual Analytics for Multivariate Time Series: The analysis of time-oriented data is an important problem for many other domains beyond movement data. In a systematic review, Aigner et al. [20] surveyed more than 100 visualization approaches and systems for time-oriented data. Many approaches for visualizing multivariate time series are based on a form of small multiples [21] where the many charts – one for each univariate time series – are juxtaposed on a common time axis. Space-efficient visualization techniques like horizon graph [22], braided graph [23], and qualizon graph [24] have been designed and experimentally evaluated for such purposes. The 'LiveRAC' sytem [25] visualizes time series for hundreds of parameters in a reorderable matrix of charts, for IT systems management. The system allows for the reordering and side-by-side comparison with different levels of detail. 'KAMAS' [26] is a knowledge-assisted visual malware analysis system, supporting IT-security experts during behavior-based malware analysis based on multivariate log files of the executed system and API calls of an operating system.

The 'PieceStack' system [27] provides an interactive environment to split and hierarchically aggregate time series based on stacked graphs [28]. 'Gnaeus' [29] provides visualizations of multivariate time series data from electronic health records using clinical guidelines for knowledge-assisted aggregation and abstraction. A different approach to tackle multivariate data applies dimensionality reduction to project multivariate measurements to 2D space, where they can be displayed as trajectories (such as a connected scatter plot) [30]. The 'TimeSeriesPaths' system [31] applied this approach as a visual data aggregation metaphor to group similar data elements. Based on their VA approach, they provided a hatching based bar visualization for inner class comparison. Schreck et al. [32] showed trajectories in small multiples and applied self-organizing maps to spatially cluster the trajectories.

The presented work focuses on multivariate time series data to solve problems in different domains. However, only one of the identified approaches (KAMAS [26]) provide the ability to extract and store implicit knowledge of experts in the form of explicit knowledge in a database. This is a desirable feature for a VA tool, especially in clinical gait analysis, as it would support clinicians in decision making when analyzing a patient's gait and would support joint learning between different clinical experts.

3 Problem Characterization & Abstraction

One primary goal in clinical decision making during gait rehabilitation is to assess whether a recorded gait measurement displays normal gait behavior or if not, which specific gait patterns (abnormalities) are present. To understand how to support the analysts in this context, we performed a 'problem characterization and abstraction', defined as the first contribution of a design study [10]. To ensure knowledgeable results for the domain of clinical gait analysis and rehabilitation, along the triangle of data, users and tasks [33], we followed a user-centered design process [34]. Information was gathered primarily from focus group meetings [35, pp. 192] and set in context with domain-specific literature. Based on this, we addressed the first (*domain problem and data characterization*) and second level (*operation and data type abstraction*) of the nested model by Munzner [11].

3.1 Focus Group

The primary aim of the focus group meetings was to match the domain-specific vocabulary between the computer scientists and clinical experts. In addition, these meetings were used to establish a mutual understanding of the following questions for the specific setting:

- What is the workflow in a clinical gait laboratory?
- How does the clinician interact within this setting?

3.1.1 Method

Participants: Seven participants comprised the focus group (two clinical gait analysis experts, two pattern recognition experts, and three visual analytics (VA) experts).

Design & Procedure: The focus group members shared a co-working space so that short stand-up meetings were possible and questions could be resolved quickly. Additionally, six focus group meetings with a duration of approximately one hour were held to discuss detailed questions with all members. All these activities were held over a 13-months time frame.

Apparatus & Materials: The results of the frequent discussions and meetings were regularly documented, which resulted in an extensive basis for a common mutual understanding. These notes were subsequently transformed into the manuscript at hand.

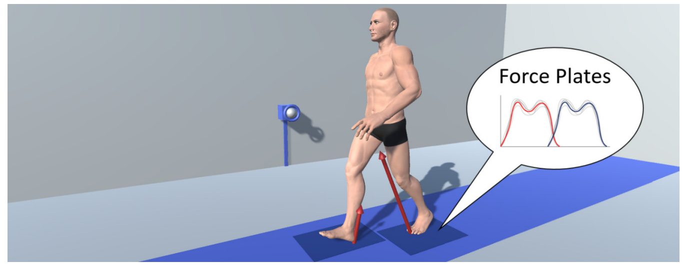


Fig. 1: During clinical gait analysis, two force plates integrated flush into the walkway allow for quantifying the ground reaction force (GRF) per foot during walking.

3.1.2 Results

A sufficient amount of patient gait data is necessary to develop visualization and pattern recognition applications for the clinical practice. While there have been attempts to provide such gait analysis databases [36], the amount of publicly available data is still too limited. Most gait data are still located directly at the clinics.

Clinical Partner during Development: The AUVA is the mandatory social insurance for occupational risks for more than 3.3 million employees and 1.4 million pupils and students in Austria and runs several rehabilitation centers. These centers typically use force plates to determine ground reaction forces (GRFs) to assess patient gait disorders and to evaluate patient progress during physical therapy treatment. This allows for a high patient throughput and accurate gait measurements. The prototype described in this manuscript was developed along the needs of the AUVA's clinical gait laboratories and clinical practice. The data incorporated in the prototype were acquired retrospectively from the AUVA's database. Thus, for the proposed developments only a subset of the available data were used.

General Gait Analysis Workflow: A typical gait analysis scenario may be divided into three main workflow stages: (1) Rigorous physical examination of the patient by the clinician: objective anatomic findings are evaluated through for example observation, palpation, and manual testing. (2) Instruction of the patient during gait analysis and data recording: the clinician guides the patient through the entire process of gait analysis and instructs the patient to walk repeatedly across an approximately 10 meter walkway (see Figure 1). In the center of the walkway, one or more force plates with an approximate size of 0.4 x 0.6 meters are fitted flush to the ground. The clinician records necessary data by operating the measurement equipment and takes care that several clean footsteps and corresponding videos are recorded. (3) Processing and interpretation of the acquired data: the clinician processes recorded data using commercial software provided by the manufacturer of the measurement equipment. These systems typically present the collected data in a non-interactive interface, as line plots of GRFs (see Figure 2) and several calculated discrete parameters (e.g., walking velocity, step length, etc.) as numbers in a table.

3.2 Data–Users–Tasks Analysis

Above we have described the general workflow of clinical gait analysis. In addition to these results, we structured the

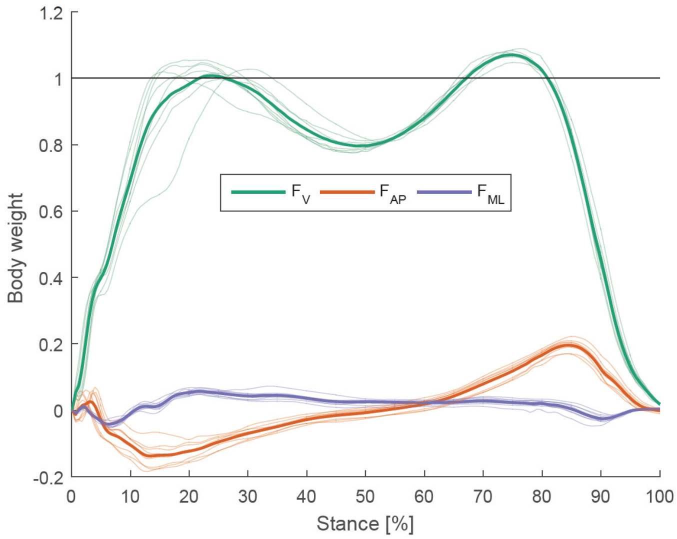


Fig. 2: Typical recordings (consistency graphs) of ground reaction force (GRF) data during a clinical gait analysis session. A total of 10 steps of the same foot are recorded, time-normalized to the stance phase of one step and amplitude-normalized to body mass. The three dimensional GRFs are presented as the vertical force component (F_V) and the anterior-posterior (F_{AP}) and medio-lateral (F_{ML}) shear forces (*the average curves are drawn in bold*).

'domain problem and data characterization' along the Data-Users-Tasks triangle [33], which will be described in the following. This high-level framework is structured around three questions:

- What kind of data are the users working with? (*data*)
- Who are the users of the VA solution(s)? (*users*)
- What are the (general) tasks of the users? (*tasks*) [33]

The answers to these questions support designers of VA methods to find or design appropriate visual representations of the data combined with appropriate analysis and interaction methods that support the users.

Data: The aforementioned force plates for gait analysis typically adopt strain gauges or piezo-electric quartz crystals arranged in force transducers to convert force into electric signals [2, pp. 311–324]. The raw signals from the transducers are amplified, digitized, and sampled with 1000 Hz. Resulting data then are processed to determine the total vertical, anterior-posterior and medio-lateral force components in Newton [2, pp. 325–326]. Even though all three components are necessary to describe a patient's gait pattern, the vertical component of the GRF receives most attention in research and clinical practice. The reason is that the vertical component shows greater magnitudes than the other two components [37]. To foster comparability of the GRF data between patients, theses three GRF components are normalized to body weight (BW) by dividing each value through the product of body mass and standard gravity (g_0). As step times vary within and between patients, data are often time-normalized to 100% stance time. Resulting data for each foot are then visualized by plotting so called consistency graphs, where all trials are plotted into one graph to inspect variability across the steps recorded (see Figure 2). Then, for visual inspection simplicity, a mean representative curve from all trials and corresponding standard deviation bands are calculated and plotted. From these data several discrete biomechanical parameters can be derived, such as amplitudes and time points of local peaks and valleys. Although GRF is a very sensitive measure of gait pathology, its specificity is low since GRF comprises the motion and acceleration of whole body dynamics [5, pp. 95]. Thus, additional measurements are necessary to describe the gait pattern of an individual in detail. In clinical gait analysis, the repeated movement of steps are referred to as a gait cycle, which starts with the initial contact of one leg with the ground, to the next ground contact of the same leg. Within this concept, one can assess spatial and temporal parameters (STPs) of gait [38]. Spatial parameters comprise the length of a step or a stride (two consecutive steps). Temporal parameters comprise the time duration of for example a single step, a stride, or the swing phase. Additionally, the cadence (steps per minute), number of gait cycles per specified time, and walking speed are used to express the temporal aspect of gait.

Typically, three GRF components from two consecutive steps are recorded during gait analysis. These data then allow to calculate additional parameters such as the STPs, which compose information about spatial (e.g., step lengths) and temporal (e.g., step time) dimensions, and rhythmical aspects (e.g., cadence) of gait.

Users: Clinical gait analysis is performed by domain experts – *physicians, physical therapists, bio-medical engineers or movement scientists*. These users have a strong background in gait and movement analysis, typically holding a university degree. They possess background knowledge about anatomy, biomechanics, gait analysis and a particular intuition on pathological gait functions. The users are comfortable using different data representations (e.g., spreadsheets, box plots, line plots), mostly developed for a special hardware setting for their institution. Thus, they have no dedicated experience with VA solutions. Generally, gait analysis is a specialist skill that requires experience, and interdisciplinary knowledge from medical and technological domains.

Tasks: The primary task of a clinician in gait rehabilitation is to *assess gait performance, to analyze and interpret the acquired data and to use this information for clinical decision making*. Secondary tasks involve the identification of specific gait patterns (abnormalities) and the comparison of observed data to already existing patient data sets (e.g., in the clinic's database). To support these tasks, expert knowledge might be stored in some sort of database, so that this information can be shared with other clinicians.

3.3 Prototype Requirements

Based on the insights gained in Section 3.2, we defined four key requirements (R) which have to be fulfilled by the KAVAGait system:

R1 Data: *Handling of complex data structures in clinical gait analysis.* To ensure the effective exploration and analysis, time-dependent multidimensional ground reaction forces (GRFs) and spatio-temporal parameters (STPs) need to be modeled, stored, and visualized for inspection. In addition, for clinical decision making a visualization of the patients' raw data is essential.

R2 Visual Representation: *Visual representations appropriate for gait analysis experts.* Clinicians use different diagrams (e.g., box and line plots) to conduct their analyses.

R3 Workflow: *Workflow-specific interaction techniques.* It is important to provide familiar interaction techniques and metaphors (e.g., drag and drop, sorting, filtering) to the clinicians, which support the identification of specific gait patterns and the comparison to already existing data sets of patients.

R4 Expert Knowledge: *Externalization of expert knowledge to reuse and share.* When analysts solve real world problems, they have a vast amount of data at their disposal to be analyzed and interpreted. By storing the clinicians' implicit knowledge, it can be made internally available in the system and usable to support the analysis process.

These four requirements form the basic pillars of KAVAGait, which have to be fulfilled during the design and implementation phase.

4 Design & Implementation

To keep the design in line with the needs and requirements defined earlier (see Section 3), we continued our user-centered design process [34] by involving three domain experts in clinical gait analysis. We iteratively produced sketches, screen prototypes, and functional prototypes [39]. Thus, we could gather and apply feedback about the design's usability and how well it supports the needs of the clinicians. This way, we addressed the third (*visual encoding and interaction design*) and fourth level (*algorithm design*) of the nested model [11]. The design study resulted in KAVAGait (see Figures 3, 5, and 6), which is implemented in Java, based on a data-oriented design [40] (e.g., used in game development and real-time rendering). Next, we elaborate on central design decisions.

4.1 Input Data

The primary input data for the KAVAGait system are the vertical component of ground reaction force (F_v) of both feet collected by force plates in the form of two synchronized time series. From these time series, spatio-temporal parameters (STPs) are calculated as 16 discrete numbers. Additional patient data on gender, age, body mass, and body height are available.

4.2 Explicit Knowledge Store (EKS)

To support gait analysts during their work, we designed the EKS related to STPs for different categories of gait patterns. Generally, the EKS stores the computerized form of the analysts' implicit knowledge as explicit knowledge. Therefore, categories of different gait patterns are used, whereby each contains 16 value ranges depending on the 16 STPs describing the gait analysis result of a patient. For each category, the EKS contains the previously assigned patients and the values of their STPs. Additionally, clinicians can refine the value range $[min, max]$ for each STP and category manually.

4.3 Visual Interface Design Concept

To best support physical therapists and gait analysis experts, we created an interface structure that allows working from *left-to-right* to fulfill the analysis tasks. This interface structure establishes an easy workflow concept based on multiple views for gait analysis experts. In relation to this interface structures, we situated the table structure of the EKS (see Figure 3:1) as well as the tree structure of the EKS (see Figures 5:1 and 6:1) to the left side of the interface to select individuals or categories of interest for exploration that always includes the related filtering options.

4.4 Visualization Concept

The KAVAGait system supports two major use cases 1) to assess newly acquired patient gait data as elicitated in Section 3 (see Figure 3) and additionally 2) to explore and adjust the stored explicit knowledge (see Figures 5 and 6).

The visual representations used in KAVAGait have been developed through continuous refinement in multiple focus group sessions. Typically, we started from sketches based on a small number of known visualization techniques for multivariate time-oriented data [20]. Those suggestions were then discussed in focus group meetings with the domain experts and continuously improved or new ones derived (e.g., ITBP) to fulfill the users' needs.

4.4.1 Assessment of Newly Loaded Patient Gait Data

When loading new 'Patient' data, the information (see Figure 3:2a) containing the 'ID', 'Age', 'Body mass', 'Body height' and 'Gender', and the measurements of the 'vertical ground reaction forces' (F_v) (see Figures 3:2b and 2c) are visualized in the center view of KAVAGait (see Figure 3:2). These F_v data are represented for each foot (see Figure 3:2b), whereby the light gray lines represent a single step and the red (left foot) or blue (right foot) line represent the mean F_v data of the single steps. Additionally, a joint representation of F_v on a combined temporal axis is available (see Figure 3:2c) for further analysis and comparison. To make the F_v of a newly loaded patient comparable with others, they are normalized by body weight.

For identification of possibly matching gait patterns, the 'Knowledge Table' (see Figure 3:1a) relates the newly loaded patient's calculated STPs to 'Categories' (pathologies) of specific gait patterns (gait abnormalities) or norm data (describing healthy gait). These 16 calculated STPs are the input for the visualizations in the 'Params in Category' column. Depending on the 16 STPs, the so called 'Graphical Summary' tells the clinician if a patient parameter x is in range $[min, max]$ with a black rectangle or if it is out of range ($x < min \vee x > max$) with a black rectangular frame based on the calculated ranges out of the EKS. If the EKS does not contain data for a category (empty category), the 'Graphical Summary' represents a gray rectangle. Thus, these three states provide a first overview of the patient. The third column ('Match') represents how the newly loaded patient matches to the stored categories in the EKS (a wider bar means a better match) supporting the clinicians during clinical decision making. For each category, a bar of width c is computed according to Equation 1:

Fig. 3: User interface of KAVAGait with its three main areas for gait analysis. 1) The table structure in (1a) shows the explicit knowledge store (EKS) that provides an overview of the stored gait patterns and how good the currently loaded data matches their definitions along with the controls in (1b) are used for filtering the included data in the EKS. 2) The patient explorer including the (2a) 'Person Information', the (2b) visualization of the 'vertical ground reaction force' (F_v) for each individual foot and the (2c) visualization of the combined F_v from both feet. 3) Shows the 'Parameter Explorer' visualizing the 16 calculated spatio-temporal parameters (STPs) of the loaded person in relation to the 'Norm Data Category' and a second 'Selected Category'.

$$c = \sum_{i=1}^{n} \frac{\sigma_i}{max(|\mu_i - x_i|^2, \epsilon)} \quad (1)$$

Equation 1 defines the matching between a sample and a 'Gait Category'. Thereby, i iterates over all 16 STPs, σ_i is the standard deviation and μ_i is the mean for the specific STP of all patients in the category. Additionally, x_i is the specific STP of the newly loaded patient. Note that Equation 1 is an inverted variant of the Fisher criterion used in linear discriminant analysis (LDA) [41]. Variable c grows with increasing agreement between x_i and the distributions defined by μ_i and σ_i. ϵ is a small number that avoids potential division by zero. By using the included filtering options (see Figure 3:1b), the explicit knowledge used for the matching calculations can be filtered by 'Gender', 'Age', 'Body height' and 'Body weight'. When selecting a category of interest, the loaded 'Patient' can be compared to other patients in the 'Parameter Explorer' view (see Figure 3:3). This table contains five columns. The first column represents the STP 'ID' to create a connection to the number in the graphical summary represented in the formerly described 'Knowledge Table'. In the second column, the 'Parameter' name is represented and column three contains the calculated STP values for the loaded patient. The fourth column provides the 'Interactive Twin Box Plot' (ITBP) (see Figure 4), an extended data visualization slider [42] for inter-category comparison in relation to (1) the 'Norm Data category' represented as blue box plot, (2) the 'Selected Category' of a specific gait patterns represented as orange box plot. A "Hatching Range Slider' (HRS) visualizing the related discrete parameters of each patient stored in the 'Selected Category' and (3) the STP value of the currently loaded patient. By placing the parts of the ITBP directly on top of each other, they can be perceived as a single control [43]. The ITBP enables the clinician to quickly compare two distributions and to set norm-value ranges for healthy and non-healthy categories. Additionally, based on the HRS, the clinician has the ability to quickly visually adjust the typical value ranges of the 'Selected Category'.

The last column represents the difference d between the 'Norm Data Category' and the 'Selected Category' which are visualized in the ITBP based on the Fisher discriminant function (see Equation 2) [41]:

$$d = \frac{(\mu_k - \mu_l)^2}{\sigma_k^2 + \sigma_l^2} \quad (2)$$

Hereby, for a given parameter, μ_k specifies the mean and σ_k^2 the variance of the first category k and respectively μ_l specifies the mean and σ_l^2 the variance of the second

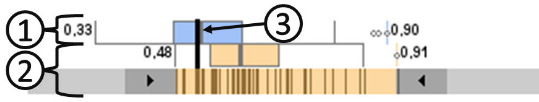


Fig. 4: Illustration of the 'Interactive Twin Box Plot' (ITBP) for intercategory comparison. 1) represents the 'Norm Data Category' as a blue box plot, 2) represents the 'Selected Category' of a specific gait abnormality as an orange box plot in combination with a 'Hatching Range Slider' (HRS) and 3) represents the actual STP values of the currently loaded patient for comparison.

category l. A higher d represents a larger difference between the parameter distributions of the two categories, yielding a wider bar. After a clinician has finished exploring the newly loaded patient data, he or she can add them to the currently selected knowledge table category in the EKS by using the 'Apply' button. This way, the parameters of the patient are automatically transferred into the EKS, recalculating the value ranges. Thereby, new explicit knowledge is generated and used for analysis support. Likewise, the clinician has the possibility to undo various changes in the EKS at any time by using the 'Reset' button.

4.4.2 EKS Exploration and Adjustment

To support its second use case, KAVAGait contains two additional views for the exploration and adjustment of the explicit knowledge stored in the EKS. The clinician has the ability to select a single 'Patient' in the EKS for comparison with other patients (see Figure 5:1). The ITBPs are showing the relation to the 'Norm Data Category' and a 'Selected Category' of abnormalities (see Figure 5:2). This visualization works the same way as formerly described for the exploration of newly loaded patient data (see Figure 3:3). On the other hand, (see Figure 6:1) the clinician can select a category visualizing each STP value range set by HRS (see Figure 6:2) included in the 'Selected Category'. Here, the clinician has the ability to change (overwrite) the automatically estimated range by moving the HRS for each parameter. This feature is needed for two specific cases: 1) If the EKS did not contain any patients in a category, the clinician has the ability to create the ranges for each STP of a category based on his or her implicit knowledge; 2) If an STP of a category contains outliers based on patient data, the clinician has the ability to readjust the range. Hereby, the color of the HRS will change to dark orange and by applying the changes, the category receives an orange triangle in the tree structure to remind the clinician that changes were applied by hand. In general, each HRS included in the visual interface (see Figures 3, 5 and 6) can be used for filtering or adjustment of the EKS included categories. The adjustments performed directly for the EKS will be stored permanently but they can be recalculated based on the values stored in the EKS if necessary. In contrast, the selected filtering options are automatically set back to default if the system is restarted.

4.5 Interaction Concept

For a better understanding of its functionality, we describe KAVAGait according to five steps based on the visual information seeking mantra by Shneiderman [44]: overview first, rearrange and filter, details-on-demand, then extract and analyze further.

Overview: When the clinician loads an input file, the patient's information and the F_v data from the performed analysis are displayed in the center of the view (see Figure 3:2). The automatically calculated matching of the patient to the stored EKS categories will be presented on the left side (see Figure 3:1a). Additionally, the 'Graphical Summary' provides an overview of the 16 represented STPs including a comparison to the stored values.

Rearrange: The clinician has the ability to rearrange each display, represented as a table, by sorting the columns (see Figures 3:1, 3:3 and 5:2).

Filter: To reduce the number of patients used for the calculation of the automated category matching, the interface offers the selection of several filtering options. Thereby, the clinician can filter the EKS data by 'Gender', 'Age', 'Body height' and 'Body mass' (e.g., see Figure 3:1b). The matching results displayed in the 'Knowledge Table' are updated immediately, and the graphical summary ('Parameters in Category') gives an impression of the 16 matched value ranges of the calculated STPs (see Figure 3:1a).

Details-on-Demand: If a matching result catches the clinician's interest, it can be selected from the knowledge table. This action opens a detailed visualization of the underlying parameters in a separate table – the 'Parameter Explorer' (see Figure 3:3), which can be also used for the exploration or already stored patients in the EKS (see Figure 5:2). In this table, the clinician can compare the calculated parameters of the loaded patient to the 'Norm Data Category' and to the 'Selected Category' from the 'Interactive Twin Box Plots' (ITBPs). Thus, the clinician still gets the information of how different the categories are for different STPs ('Category Difference'). It is important to note that the ITBPs are situated above each other to provide a better visual comparability of the differences between the *left and the right foot*. This is important, as the clinician needs to assess differences between both body sides (gait asymmetry). Additionally, the clinician has the ability to sort the visualized data based on the different columns by clicking the respective header.

Extract: Once the clinician has found the appropriate category for a patient's gait, the calculated parameters can be added to the 'Selected Category' of the EKS by pressing the 'Apply' button (see Figure 3:3). Alternatively, the clinician can select some parameters of the patient in the 'Parameter Explorer' table to add only them to the 'Selected Category' of the EKS by using the 'Apply' button. From this moment, these data are immediately integrated into the automated analysis for the matching calculation. If a class contains insufficient samples or a value range is affected by outliers, the clinician has the possibility to extract further implicit knowledge by manually adapting these ranges in the 'Category Parameter Explorer'. For this purpose, the ITBPs utilize the raw data of the selected class in order to provide the possibility for visual control by the clinicians (see Figure 6:2).

4.6 Externalized Knowledge Integration

The EKS is included on the left side of the interface in two different forms depending on the task to be supported.

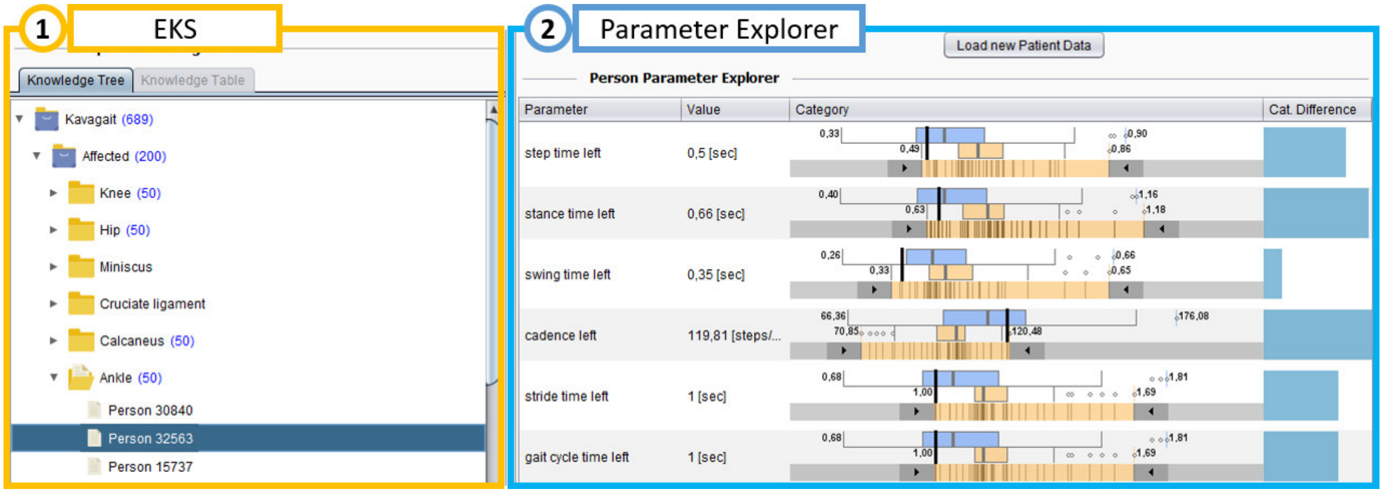


Fig. 5: User interface for 'explicit knowledge store' (EKS) exploration and adjustment in relation to stored single patients. 1) The tree structure of the EKS while selecting a single patient for comparison and adjustment 2) with other patients in relation to the norm data category and the category that includes the patient (Ankle in this case).

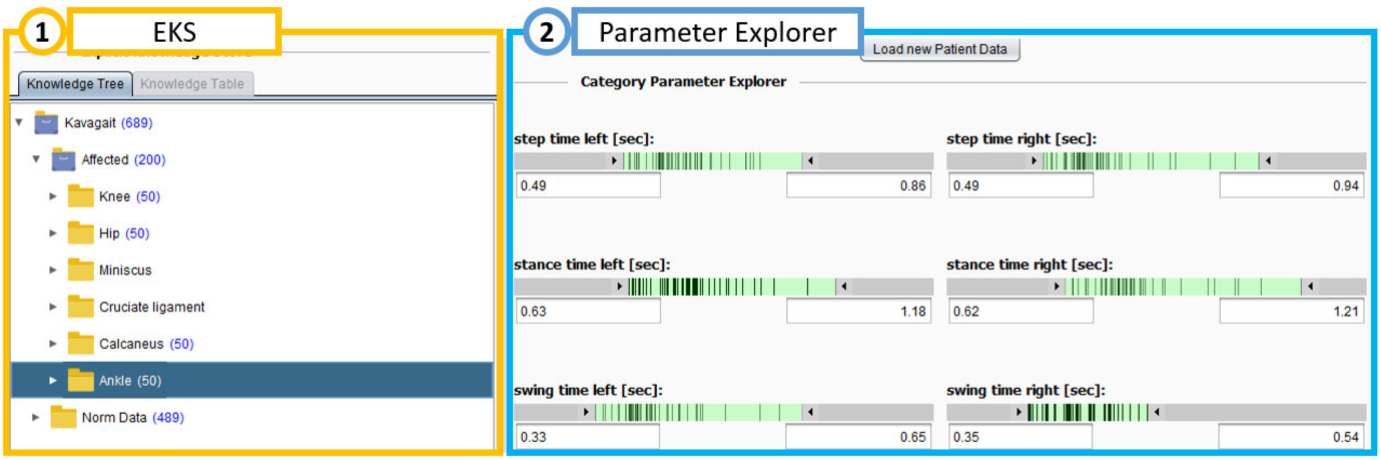


Fig. 6: User interface for 'explicit knowledge store' (EKS) exploration and adjustment in relation to categories containing several patients. 1) The tree structure of the EKS while selecting a category ('Ankle' in this case) for comparison 2) based on the 'Hatching Range Slider' (HRS) representing the limits or norm-ranges of the spatio-temporal parameters from the patients included in the category.

First, when the clinician explores newly loaded patient data, the EKS is presented in a table format (the 'Knowledge Table') (see Figure 3:1.a). All of the 'Categories' which are integrated in the EKS will be checked against the loaded input data automatically. Based on the explicit knowledge, the system distinguishes between the three states (in range, out of range, or no data) of the graphical summary for each of the 16 STPs in the 'Parameters in Category' column. Additionally, the system calculates how newly loaded patients match to the stored 'Categories' in the EKS. To add new knowledge to the EKS, the system provides two possibilities: On the one hand, the clinician can add the full patient dataset, representing each parameter as ITBP, by using the 'Apply' button in the 'Parameter Explorer' (see Figure 3:3) to the 'Selected Category' in the 'Knowledge Table'. On the other hand, the user has the ability to select a set of parameters of interest from the 'Parameter Explorer' table and add them using the 'Apply' button in the 'Parameter Explorer' to the 'Selected Category'.

Secondly, when the clinician explores the explicit knowledge, and adjusts it for single patient data or a category, the EKS is presented as an indented list (the 'Knowledge Tree') (see Figure 5:1). On the one hand, (see Figure 5:1) the clinician has the ability to select a single 'Patient' from the EKS for comparison (see Figure 5:2) with other patients by using the ITBP in relation to the 'Norm Data Category' and the 'Selected Category' including the selected 'Patient'. On the other hand, (see Figure 6:1) the clinician can select a category visualized by HRS (see Figure 6:2) for each STP of the patients included in the 'Selected Category' of the EKS. Generally, at the end of each 'Category', the number of contained 'Patients' is shown in blue brackets.

Knowledge Generation Loop: Figure 7 provides an overview of the system's knowledge generation loop, starting at the dark gray inner loop. In general, the system's EKS stores all 'Patient' data in several 'Categories' depending on patients' pathologies (gait abnormalities) which were generated by former gait analysis sessions. If the clinician

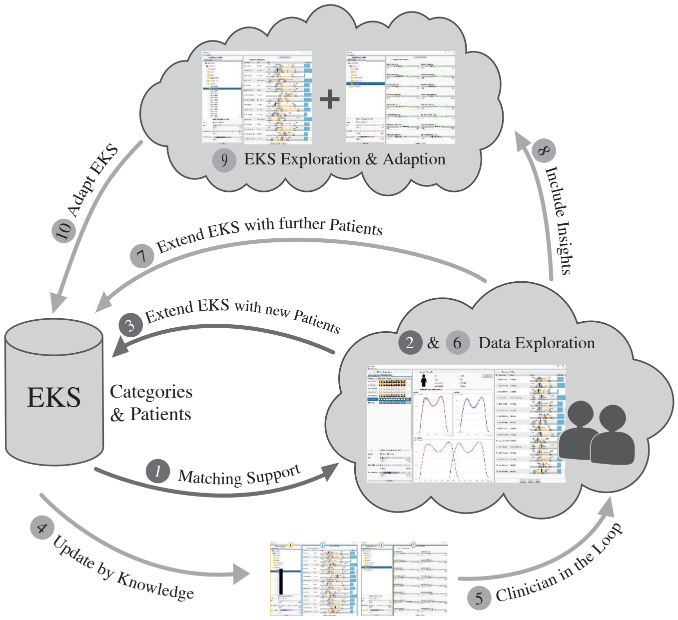


Fig. 7: Overview of the system's knowledge generation loop, beginning with the dark gray inner loop after loading a new patient and changing to the light gray outer loops for interactive data exploration and knowledge generation. The clinician plays a major role in both loops.

loads a new patient file, the calculated STPs will be checked automatically against the EKS (1) to calculate the category matching. Depending on the automated matching calculations, the system provides a visual representation of the results (2). From this point, the clinician can carry out the patient data exploration and analysis, thus, the clinician is an important part of the knowledge generation loop. (3) During the patient data analysis driven by the clinician, the clinician has the ability to include the patient data into a 'Category' in the EKS. By adding a new 'Patient' to the EKS or setting filters, the system automatically refreshes the matching calculations depending on the explicit knowledge stored in the system (4). This brings the user into the outer (light gray) part of knowledge generation loop. Here the clinician is part of the continuously recurring loop (5), for data exploration (6) and knowledge generation (7). Additionally, the clinician has the ability to continuously include new insights (i.e., gait categorizations, value range limits) (8) depending on 'Patient' and 'Category' exploration and adjustment (9) to adapt the EKS value ranges, which are predefined by the stored explicit knowledge, for further automated analysis (10).

5 Validation & Results

To validate the KAVAGait system and provide evidence for its effectiveness, we followed a threefold research approach consisting of moderated expert reviews, user studies [35] and a case study with a national expert. All of the insights were documented in detail to ensure reproducibility [45] and used to improve our research prototype. All materials used, such as interview guidelines and tasks, are available as supplemental material (https://phaidra.fhstp.ac.at/detail_object/o:1928).

5.1 Expert Reviews

In the first step, we conducted iterative expert reviews to eliminate usability issues in the basic functionality and appearance of the interface.

5.1.1 Method

Participants: To validate the visual interface design, we invited three usability experts. Each of them has between two to four years of experience in this field. Two of them are between 20 and 29 years of age and one is between 30 and 39 years of age. All of them have a Master's degree and advanced or expert knowledge in usability.

Design and Procedure: Each usability expert received a short introduction to the basic features and the workflow of the system. Next, each expert walked through each feature individually and assessed usability issues.

Apparatus and Materials: As evaluation material, we generated builds of KAVAGait in different development states and used them for the iterative expert review sessions performed on a 15″ notebook with a full HD screen. Each expert review was documented in short notes on paper by the investigator.

5.1.2 Results

The basic color scheme of KAVAGait was found to be easily recognizable. Only the coloring of the 'Categories' was pointed out as being not well differentiated from the other elements (see Figure 3). The visualization metaphors (boxes, folders and sheets) for the knowledge tree visualization was developed in conjunction with the usability experts to represent a familiar structure to the analysts. The experts suggested that it is necessary that the interface automatically applies the entered parameter if the user left the focus of a filtering input box. Overall, all of the usability experts provided positive feedback on the design structure of the system. All of the expert's suggestions were used for a redesign and revision of the system in order to prevent the domain users from having basic interface issues.

5.2 User Study

A user study with six gait analysis experts was performed in October 2016 as formative evaluation of usability [46] on the revised system. Each test took approximately 1.5 hours and encompassed four analysis tasks, the system usability scale questionnaire (SUS) [47], and a semi-structured interview built upon 13 main questions to be answered. The user study's goals (G) and non-goals (NG) are defined as: (G1) Testing the functionality of the research prototype; (G2) Testing the visualization techniques for comprehensibility in relation to the domain; (G3) Testing the utility of knowledge storage and representation in the system; (NG1) Comparison of KAVAGait with another analysis system and (NG2) conducting performance tests, because there was no comparable interactive analysis system found for this domain.

5.2.1 Method

Participants: We invited six gait analysis experts (see Table 1) to participate in the user study. One participant had given feedback on sketches and early prototypes previously as a member of the focus group for the user-centered design process (see Section 4). All experts work in the field of clinical gait analysis as physical therapists, bio-medical engineers or sports scientists. Additionally, all of them are involved in different gait analysis or physical therapy research projects and two of them are also working as physical therapist in a hospital.

TABLE 1: Data on user study participants describing education, years in the field and knowledge in gait analysis. (Gender: f := female, m := male; Organization: R := research, F := faculty, H := hospital; Knowledge: 1 := basic, 2 := skilled, 3 := advanced, 4 := expert)

Person (Gender)	Organization	Age	Knowledge	Years in Field	Education
P1 (f)	F	40-49	2	15	MSc
P2 (m)	F	40-49	3	15	PhD
P3 (m)	R&F	30-39	2	5	PhD
P4 (f)	R	40-49	1	1.5	MSc
P5 (f)	F&H	30-39	2	10	MSc
P6 (f)	F&H	20-29	2	7	MSc

Design and Procedure: At the beginning of the user study, each participant was asked about a general impression of the user interface and which functions could be recognized. This step took approximately five minutes. Subsequently, each participant had to solve **four guided analysis tasks**. The first three included a step-wise introduction to the system and the last one was a combined analysis task to determine the understanding of the system's workflow (this step was also required for the subsequent SUS questionnaire). Each analysis task was read to the participant at the beginning of the task, and for reference, each task was handed over to the participant in printed form. For the analysis tasks, participants spent approximately 40 minutes. After the analysis task session, each participant had to fill out a standardized **SUS** questionnaire [47] in less than five minutes. Finally, we performed **semi-structured interview** sessions with an average duration of 40 minutes. For this, we used an interview guideline consisting of 13 major questions addressing general system usability, filtering, using the 'explicit knowledge store' (EKS), and individual visual metaphors used in KAVAGait.

Apparatus and Materials: The user studies were performed in a silent room and the **analysis tasks** were performed on a 15″ notebook with full HD screen resolution and an external mouse. As datasets for the analysis tasks, we used two anonymous clinical gait analysis samples recorded by an AUVA clinical gait laboratory. The provided datasets included one healthy and one patient with a gait abnormality. To achieve the best outcome, we asked the participants to apply thinking aloud [48] during the whole analysis task part. For further analysis of the user test, we recorded the screen and the participant using the notebook's internal webcam. In parallel, the facilitator took notes in our predefined test guideline. The **SUS** questionnaire and **semi-structured interview** were conducted on paper in the participants' native language. For the detailed questions, we used small images in the semi-structured interview guidelines to support the participants in recalling the respective parts of KAVAGait.

5.2.2 Results

The following section describes the major findings of each part of the test and summarizes their results.

Analysis Tasks: During the four analysis tasks, all participants described their solutions to the tasks at hand by thinking aloud. They expressed problems as well as benefits of KAVAGait during their work. One major problem during the test was that the participants tried to find out which rectangle in the graphical summary, displayed in the knowledge table, relates to which 'Interactive Twin Box Plot' (ITBP) in the parameter explorer [P2, P3, P5, P6]. In relation to the graphical summary, all participants understood the gray rectangle, the empty black frame and the black rectangle shape encoding the matching of the individual parameters in relation to 'no data available', 'out of range' and 'is in range'. Four out of six participants stated during the third analysis task that they confirm the automated matching result after comparing to the vertical ground reaction force (F_v) data. In general, all participants stated that the automated matching can be used as a guided entry point for further analysis using the twin box plots to analyze the patients' data in detail. Additionally, all participants handled the tasks depending on the knowledge exploration or own knowledge externalization very well. They understood the knowledge tree organization metaphors (boxes, folders and sheets) visualizing categories and individual persons.

System Usability Scale (SUS): By applying the SUS, we were able to obtain a good indication concerning the handling of our KAVAGait system. The results show a SUS value of 80 points in average out of 100, which can be interpreted as good without significant usability issues according to the SUS description by Bangor et al. [49]. Comparing 500 SUS scores, Sauro [50] found that only 10% of systems reached an SUS score greater than 80.

Semi-structured Interviews: Next, we present the results of the performed interviews, which were structured along 13 main questions. The detailed evaluation of the interviews can be found in our supplement material (https://phaidra.fhstp.ac.at/detail_object/o:1928).

All participants attending the user study confirmed a very clear system design. The used visualization metaphors are described as well interpretable, the color scheme is conclusive and the calculated matching acts trustworthy. The integrated filter options for data reduction make sense and are understandable. The various visualization options included in KAVAGait and the curves and box plots contributed very well and were considered understandable. Only one participant, P6, indicated that it was not clear how the 'Category Difference' was calculated.

Generally, all participants told us that they readily understood the EKS, including the ability to compare newly loaded patients to the EKS for categorization. Additionally, a single system would be very beneficial as well as a shared EKS. The knowledge representation options ('Knowledge Tree', 'Knowledge Table', 'Hatching Range Slider' (HRS) and ITBP) were described as very helpful and good for quick decision making. They described that it was easy to get an overview of the loaded patient in relation to the 'Norm Data Category' and the 'Selected Category' based on the ITBPs. Thus, the analyst has the ability to navigate through the data or to rely on the EKS based matches.

Additionally, the participants referred to the categories of the EKS, including patient data of prior analysis, as well as to the different representations in the 'Knowledge Table'. Four out of six participants reported that based on the ITBP the sharp distinction, interval range, variance, differences and relationships can be derived for the analysis.

All participants noted that they have understood the saving process to the EKS. The category in which a patient is stored should appear subsequently. As 'nice to haves', a separated saving area for not yet diagnosed patients and the ability for annotations would be helpful. The separated category should not be included in the automated matching calculations. Furthermore, all participants reported that they understood the symbols represented in the 'Knowledge Tree' of the EKS but most of them did not assign any further meaning to them. Based on the different colors it was easy for the participants to distinguish the different EKS levels.

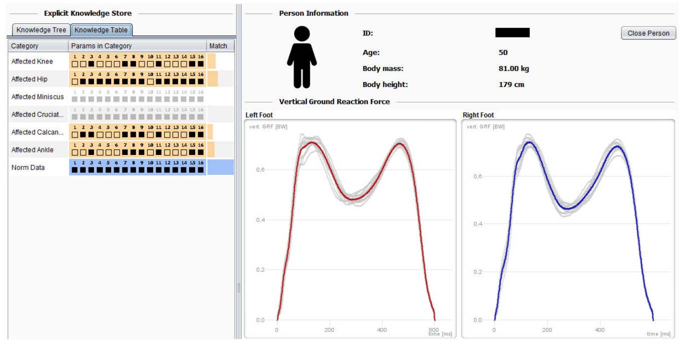
In relation to the 'Knowledge Table', the participants classified the 'Graphical Summary' as very helpful. Only two participants had small issues at the beginning when interpreting this visualization metaphor. Based on the coloring and the boxes for the parameters it was easy to understand the matching results. The matching criterion was indicated as helpful and good starting and orientation point for analysis. Some participants argued that a connection based on sequence numbers between the 'Graphical Summary' and the ITBPs would be very helpful. This suggestion and other suggestions were subsequently implemented as shown in Figure 3. The range slider shading and its usability for range assessment was quickly recognized by most of the participants (e.g., used for filtering, category parameter exploration and adjustment). The shading gave an overview of the underlying data derivation and outliers. Additionally, it helps to explain the ITBPs shape.

All participants had prior exposure to statistical box plots and therefore readily understood the ITBPs. They noted that this metaphor contains a lot of information but it is clearly structured. Additionally, some participants stated that the usage of the same color will be better understandable when selecting the 'Norm Data Category' for comparison in the ITBP (see Section 4) in the future. The comparison of a (newly loaded) patient with the 'Norm Data Category' and a 'Selected Category' based on the ITBPs was described as helpful. It is possible to see the differences of the individual category parameter, but you have to know how to interpret them.

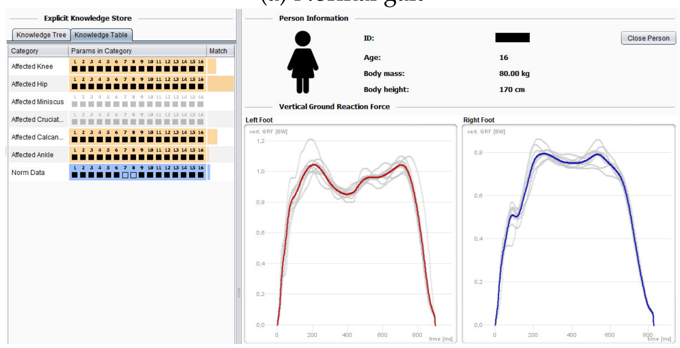
In general, the KAVAGait system was described as very innovative and helpful for the analysts by providing automated analysis and pointing out possible reasons to be respected in clinical decision making.

5.3 Case Study

After finalizing the expert review and the user study we made the following improvements before conducting the final case study: 1) ID numbers for parameter identification were added to the ITBP and the 'Graphical Summary'; 2) The coloring scheme of the ITBP was improved; 3) Tool tips were added to all elements; 4) The entire labeling used in the system was checked for consistency.



(a) Normal gait



(b) Hip associated gait abnormality

Fig. 8: Examples of the gait analysis data sets presented and discussed during the case study containing patterns of (a) normal gait and (b) gait abnormality.

5.3.1 *Method*

Participants: For our case study, we invited one leading national expert for gait rehabilitation to test and comment on our novel KAVAGait system. The expert has more than one decade of experience in conducting clinical gait analysis. Thus, the expert is comfortable in identifying gait patterns based on the representation of ground reaction forces (GRFs) and the calculated spatio-temporal parameters (STPs).

Design and Procedure: At first, the expert received a short introduction, in the form of a presentation, about the basic features and workflow of the system. Next, the expert walked through each feature individually by an example and was asked to critically comment on the system. Additionally, the expert could choose different patients from our data pool to explore them and tell us new insights gained using KAVAGait.

Apparatus and Materials: We met in the expert's office room to perform our case study. As materials, we used a short presentation of the KAVAGait system and a build of the revised prototype including 489 anonymized patients in the 'Norm Data Category' and 50 patients for each out of four patient categories (ankle, calcaneus, hip and knee associated gait abnormalities). The case study was performed in the same setup as in the user study before. The suggestions and comments stated by the expert were documented by the presenter and one observer.

5.3.2 *Results*

The expert initially noted that a clinician normally focuses on two major aspects. Firstly, they look for asymmetries comparing the STPs between the left and the right foot and secondly for deviations of them with respect to the norm data.

EKS Patient Exploration: Next, the expert randomly selected a patient stored in the 'Norm Data Category'. Generally, the expert found an added value in relation to the ITBP representation for the STPs. A key statement by the expert was: "You do not have to hit the median value directly, it is more important if the parameter values for the left and the right foot are looking similar" near to the median value of the 'Norm Data Category'. By randomly selecting a patient from the 'Knee Category' for exploration, the expert was immediately able to sort the STPs. An additional suggestion was to provide the ability to select two groups for comparison. In this case, for example, the expert could compare the 'Knee Category' to the 'Ankle Category' without changing by viewing the related mean curves of each category with their associated one-standard deviation bands. Additionally, a more detailed separation of the datasets (e.g., for abnormalities in the left, the right or both feet) in the EKS was stated as helpful to activate and deactivate these groups directly in the future.

New Patient Data Exploration: To test the analysis abilities of KAVAGait, the expert successively loaded four clinical gait analysis records, which were containing gait abnormalities (1x norm, 2x hip, 1x ankle). Two of these examples are illustrated in Figure 8. The representation of the individual steps represented as light gray curves in the background with their corresponding mean value curve was described as very helpful. The expert analyzed each loaded patient file exactly with regard to the F_v, the calculated parameters represented in ITBPs, and the system-internal computed matching based on the EKS. During the case study, several of the assessed records indicated multiple matching to different gait patterns (which were reported in the 'Knowledge Table'). The expert was able to confirm the EKS based matching results, and reported that the system was of valuable support during gait record examination. This can be attributed to the fact that the patients usually undergo a therapy for a specific problem. However, KAVAGait also recognizes abnormalities in other joints caused by specific problems. The expert stated that the community has always been aware that one abnormality can lead to others, but it has never been so well presented.

Concluding Discussion: To date the system only incorporates a total of approximately 500 patients. Thus, in each category the chance for imbalances between sub-groups, such as defined by gender, age, body mass, or height may be present. These in consequence, may introduce a bias when calculating the final matching. Future prospects may include a direct connection the AUVA database to include a sufficient and representative number of patients per class to overcome this problem.

The expert also noted, that a further development could be to visualize and store additionally derived and commonly used discrete parameters (next to the STPs) from the entire GRF curves, such as local minima and maxima, loading rates, among many others. This would further strengthen the system's capabilities of describing a patient's gait performance. Another feature, noted to be of valuable interest, would be the possibility to compare a patient with earlier treatments. This would help to visualize the entire rehabilitation process over time.

In general, the expert outlined that the system supports the process of gait assessment in several ways: 1) The system easily allows to compare new patient data to stored EKS data, and thus helps to gather a more conclusive picture of the patient's gait performance, which was not possible in clinical practice before. 2) The matching criteria of KAVAGait helps to clearly visualize and identify secondary gait impairments. This is very important, as during clinical examination secondary gait impairments are easily overlooked. 3) The ability to compare the norm data with a selected category for direct comparison based on the ITBP is very helpful to gain further insights into the patient's gait performance and possible impairments. 4) The KAVAGait system is well suited for educational training as well as for analysts who do not perform clinical gait analysis every day. For experts who are working every day in this area, the system can be a good supplement during the analysis process. Overall, the expert described KAVAGait as an excellent and helpful analysis tool.

6 Reflection & Conclusion

Following the design study methodology of Sedlmair et al. [10], the reflection (including retrospective analysis and lessons learned) is the third contribution of a design study which enables the improvement of current guidelines. The following paragraphs describe the reflection in line with the initially stated requirements from Section 3.3 (R1 – R4).

R1 Data: The data structure resulting from force plate recordings, are synchronized time series data of the ground reaction forces (GRF). Especially in a clinical gait analysis setting, these data comprise time series of two force plates (one per foot). Based on these data, we calculated eight additional STPs (spatio-time parameters) parameters, which were used for automated and visual comparison. On the one hand, we designed the 'Graphical Summary' for data comparison, showing the analyst how a single patient parameter is related to the parameter set of a category. On the other hand, we designed the Ìnteractive Twin Box Plot (ITBP) for detailed inter-category parameter comparison between the 'Norm Data Category', a 'Selected Category' of a specific gait pattern and the patient to analyze.

R2 Visual Representation: In general, the decision for an interface providing a clear structure and understandable visual representations was well received. It was easy for the domain experts in our validation to understand the handling and to work with it. Additionally, they appreciated the prototype's wide range of coordinated features, which they regarded as useful for data exploration and analysis while not being overloaded. A particularly interesting outcome of the tests from the visualization design perspective was, that the HRS are very useful for a parameter overview and range adjustments during patient data exploration as well as for the 'explicit knowledge store' (EKS) exploration and adaption. Additionally, the ITBPs were considered as well-suited for intercategory comparison in relation to a single patient parameter. This way, it is possible to see how well the patient develops in the direction of the 'Norm Data Category' for example. Another particularly notable outcome is that the 'Graphical Summary' and the 'Matching Criteria', represented in the 'Knowledge Table', were described as very valuable by the national gait rehabilitation expert. To date the system only provides information of the vertical GRF component. To increase the quality of the analysis, in the future we will extend the visualization to all three force components of the GRFs. In addition, we will also add discrete key GRF parameters of those components to further improve automated analysis and exploration. Additionally, direct comparisons of left and right foot (i.e., gait asymmetry indices [51]) need to be integrated in the twin box plots.

R3 Workflow: All included filter methods and the dynamic query concept providing a very fast response in KAVAGait were very well received. In general, the participants described the filtering, analysis, and exploration abilities as intuitive and the usage and adaption of the EKS as easy to use. Our national expert mentioned that KAVAGait is very valuable and suits several use cases, such as support for clinical experts, assistance for less experienced clinicians, and learning and training opportunities for students. Based on the insights we gained during our validation studies we found that for the participants and the national expert, the visual representations of the expert knowledge and the handling of the EKS was simple to understand and to use. To further improve the workflow, the ability to annotate the patients' data would be a helpful feature.

R4 Expert Knowledge: As previously mentioned, the 'Knowledge Tree' and the 'Knowledge Table' of the EKS were well received by the participants and the case study member. The knowledge organization as boxes, folders (categories) and sheets (patients) was well received by most of the participants. Based on the counter after each category (see Figures 3:1 and 3:3), it was easy to understand how meaningful the data are for comparison and to get an overview of the included data. As a future step, explicit knowledge should also be used in the visual analytics (VA) workflow to train machine learning methods for improved automated categorization.

Lessons Learned: As described in Section 3, clinicians currently are using non-interactive line plots and tables. For clinical decision making they are using their implicit knowledge based on several years of experience. During this design study, we learned that explicit knowledge extracted from the clinicians implicit knowledge opens the possibility to support clinicians during clinical decision making. Additionally, KAVAGait could also be used to share the knowledge of domain experts and for educational support.

For keeping up with the large number of patients stored in the EKS, clinical gait analysts need to continuously adapt the systems settings during the clinical decision making process. Supporting such interactive workflows is a key strength of visualization systems. Clinical gait analysis in particular profits from extensive interaction and annotation because it is a very knowledge-intensive job. By providing knowledge-oriented interactions, externalized knowledge can subsequently be used in the analysis process to support the clinicians. Our newly developed visual metaphors provide an easy way to inspect variability of the data (e.g., standard deviation), allow to identify outliers in the data, and provide an easy to understand overview of the data and automated matching results (as demonstrated in Figure 3:1a). Additionally, based on the ITBPs (see Figure 4) it is possible to perform intercategory and patient comparisons by details on demand to find similarities in the data.

Limitations & Future Directions: KAVAGait is a design study investigating how interactive knowledge-assisted VA methods can aid clinical decision making in the context of gait analysis. Since the system is still a proof of concept, some limitations exist. These, however, point out future directions of research in both areas, VA and clinical gait analysis. Currently, the proposed system only incorporates the vertical ground reaction force component, as used by several studies [52], [53]. Nevertheless, it is subject to future work to include the other force components as well. This will help the clinicians to get a more holistic view of one's gait performance and will further strengthen the system's capability in supporting clinical practice. Another limitation might be associated with the defined parameters for matching and comparing purpose. To date only a set of the most commonly used STPs are included. However, there are several other discrete parameters that could be used for analysis and matching purpose. Research has shown that sophisticated machine learning algorithms bear the potential to identify and cluster gait patterns [54], [55]. For example automated patient categorization based on unsupervised (e.g., [56]) and supervised (e.g., [57]) approaches might be interesting. Both aspects will support in drawing more precise medical decisions based on the data available. Results, however, clearly state, that the entire waveforms as input variables result in higher classification accuracies than using discrete parameterization techniques. Thus, future work might opt to include both, discrete parameters to inform clinicians and machine learning techniques, which use the entire waveforms, to allow for more advanced pattern recognition abilities and classification functionalities. Another future direction might be to provide the ability for searching the most similar dataset stored in the EKS to the currently viewed patient data. At this time, such a mechanism is not included in KAVAGait. Currently, KAVAGait offers the ability for comparison of each parameter for both feet of one patient. Gait symmetry plays a key role for clinicians in analyzing and interpreting gait data. To date KAVAGait only allows for visual inspection of parameters of the left and right body side. To increase the quality of such comparisons, parameters such as the 'gait asymmetry index' (GAI) might be valuable and should be included in future versions (e.g., [58]). Finally, the clinical decision support provided by KAVAGait needs to be evaluated for the effect of cognitive biases such as the confirmation bias [59] that might increase due to previously externalized knowledge and interactive steering. KAVAGait addresses such concerns by prominently providing clinicans with raw GRFs displayed as curves. Our interviews indicate that these GRF curves are always considered in decision making. A further research direction is also the integration of information on provenance and certainty into the EKS.

In contrast to existing gait analysis systems that often come with gait analysis hardware, KAVAGait enables: interaction; a knowledge store; comparison with stored data, i.e., hundreds of patients effectively; and interactive filtering. Therefore, KAVAGait uses analytical and visual representation methods to provide a scalable and problem-tailored visualization solution following the VA agenda [6], [7]. The knowledge generation loop (see Figure 7) can be generalized for other domains taking into account domain-specific data structures and patterns of interest. On a general level, this workflow for knowledge generation and extraction always includes the user as an integral part of the loop [60].

Acknowledgments

This work was supported by the Austrian Science Fund (FWF) via the "KAVA-Time" project (P25489) and by the NFB – Lower Austrian Research and Education Company and the Provincial Government of Lower Austria, Department of Science and Research ("IntelliGait" LSC14-005). Cordial thanks to Marianne Worisch, Christina Niederer and Niklas Thür for their support as well as to Tarique Siragy for proofreading our manuscript. We would also like to thank all focus group members and test participants who have agreed to volunteer in this project.

References

[1] W. Erickson, C. Lee, and S. von Schrader, "2014 Disability Status Report: United States," Cornell University Yang Tan Institute of Employment and Disability (YTI), Ithaca, NY, Tech. Rep., 2016. [Online]. Available: http://www.disabilitystatistics.org/
[2] B. Nigg and W. Herzog, *Biomechanics of the musculo-skeletal system*, 3rd ed. Hoboken NJ: Wiley, 2007.
[3] D. A. Winter, *Biomechanics and motor control of human movement*, 3rd ed. Hoboken NJ: Wiley, 2005.
[4] A. Cappozzo, U. Della Croce, A. Leardini, and L. Chiari, "Human movement analysis using stereophotogrammetry: Part 1. theoretical background," *Gait & Posture*, vol. 21, no. 2, pp. 186–196, 2005.
[5] C. Kirtley, *Clinical gait analysis theory and practice*. Edinburgh: Elsevier, 2006.
[6] J. J. Thomas and K. A. Cook, Eds., *Illuminating the path: The research and development agenda for visual analytics*. IEEE Comp. Society Press, 2005.
[7] D. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann, Eds., *Mastering the information age: solving problems with visual analytics*. Goslar, Germany: EG, 2010.
[8] M. Chen, D. Ebert, H. Hagen, R. Laramee, R. Van Liere, K.-L. Ma, W. Ribarsky, G. Scheuermann, and D. Silver, "Data, information, and knowledge in visualization," *CG&A*, vol. 29, no. 1, pp. 12–19, 2009.
[9] X. Wang, D. H. Jeong, W. Dou, S.-W. Lee, W. Ribarsky, and R. Chang, "Defining and applying knowledge conversion processes to a visual analytics system," *C&G*, vol. 33, no. 5, pp. 616–623, 2009.
[10] M. Sedlmair, M. Meyer, and T. Munzner, "Design study methodology: Reflections from the trenches and the stacks," *TVCG*, vol. 18, no. 12, pp. 2431–2440, 2012.
[11] T. Munzner, "A nested model for visualization design and validation," *TVCG*, vol. 15, no. 6, pp. 921–928, 2009.
[12] G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel, *Visual analytics of movement*. Berlin: Springer, 2013.
[13] C. Perin, R. Vuillemot, and J. D. Fekete, "SoccerStories: A kick-off for visual soccer analysis," *TVCG*, vol. 19, no. 12, pp. 2506–2515, Dec. 2013.
[14] H. Janetzko, D. Sacha, M. Stein, T. Schreck, D. A. Keim, and O. Deussen, "Feature-driven visual analytics of soccer data," in *IEEE VAST*, Oct. 2014, pp. 13–22.
[15] A. Vögele, B. Krüger, and R. Klein, "Efficient Unsupervised Temporal Segmentation of Human Motion," in *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, ser. SCA '14. Aire-la-Ville, Switzerland: EG, 2014, pp. 167–176.
[16] J. Bernard, N. Wilhelm, B. Krüger, T. May, T. Schreck, and J. Kohlhammer, "MotionExplorer: Exploratory Search in Human Motion Capture Data Based on Hierarchical Aggregation," *TVCG*, vol. 19, no. 12, pp. 2257–2266, Dec. 2013.
[17] S. Jang, N. Elmqvist, and K. Ramani, "Motionflow: Visual abstraction and aggregation of sequential patterns in human motion tracking data," *TVCG*, vol. 22, no. 1, pp. 21–30, Jan 2016.
[18] O. Purwantiningsih, A. Sallaberry, S. Andary, A. Seilles, and J. Azé, "Visual analysis of body movement in serious games for healthcare," in *IEEE Pacific Visualization Symposium (PacificVis)*, Apr. 2016, pp. 229–233.
[19] N. Wilhelm, A. Vögele, R. Zsoldos, T. Licka, B. Krüger, and J. Bernard, "FuryExplorer: Visual-Interactive Exploration of Horse Motion Capture Data," in *Visualization and Data Analysis (VDA 2015)*, Feb. 2015.
[20] W. Aigner, S. Miksch, H. Schumann, and C. Tominski, *Visualization of Time-Oriented Data*. London: Springer, 2011.
[21] E. R. Tufte, *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press, 1983.
[22] J. Heer, N. Kong, and M. Agrawala, "Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations," in *SIGCHI*. ACM, 2009, pp. 1303–1312.
[23] W. Javed, B. McDonnel, and N. Elmqvist, "Graphical perception of multiple time series," *TVCG*, vol. 16, no. 6, pp. 927–934, Dec. 2010.
[24] P. Federico, S. Hoffmann, A. Rind, W. Aigner, and S. Miksch, "Qualizon Graphs: Space-efficient time-series visualization with qualitative abstractions," in *Proc. 2014 Int. Working Conf. Advanced Visual Interfaces, AVI*. ACM, 2014, pp. 273–280.
[25] P. McLachlan, T. Munzner, E. Koutsofios, and S. North, "LiveRAC: Interactive visual exploration of system management time-series data," in *SIGCHI*. ACM, 2008, pp. 1483–1492.
[26] M. Wagner, A. Rind, N. Thür, and W. Aigner, "A knowledge-assisted visual malware analysis system: design, validation, and reflection of KAMAS," *COSE*, vol. 67, pp. 1–15, 2017.
[27] T. Wu, Y. Wu, C. Shi, H. Qu, and W. Cui, "PieceStack: Toward better understanding of stacked graphs," *TVCG*, vol. 22, no. 6, pp. 1640–1651, Jun. 2016.
[28] S. Havre, E. Hetzler, P. Whitney, and L. Nowell, "ThemeRiver: Visualizing thematic changes in large document collections," *TVCG*, vol. 8, no. 1, pp. 9–20, 2002.
[29] P. Federico, J. Unger, A. Amor-Amorós, L. Sacchi, D. Klimov, and S. Miksch, "Gnaeus: Utilizing Clinical Guidelines for Knowledge-assisted Visualisation of EHR Cohorts," in *Proceedings of the EuroVis Workshop on Visual Analytic, EuroVA*, E. Bertini and J. C. Roberts, Eds. EG, 2015.
[30] S. Haroz, R. Kosara, and S. Franconeri, "The connected scatterplot for presenting paired time series," *TVCG*, vol. 22, no. 9, pp. 2174–2186, 2016.
[31] J. Bernard, N. Wilhelm, M. Scherer, T. May, and T. Schreck, "TimeSeriesPaths: Projection-based explorative analysis of multivariate time series data," *JWSCG20*, vol. 20, no. 2, pp. 97–106, 2012.
[32] T. Schreck, J. Bernard, T. v. Landesberger, and J. Kohlhammer, "Visual cluster analysis of trajectory data with interactive Kohonen maps," *IVS*, vol. 8, no. 1, pp. 14–29, Mar. 2009.
[33] S. Miksch and W. Aigner, "A matter of time: Applying a data-users-tasks design triangle to visual analytics of time-oriented data," *C&G*, vol. 38, pp. 286–290, 2014.
[34] H. Sharp, Y. Rogers, and J. Preece, *Interaction Design: Beyond Human-Computer Interaction*, 2nd ed. Chichester: Wiley, 2007.
[35] J. Lazar, J. H. Feng, and H. Hochheiser, *Research Methods in Human-Computer Interaction*, 1st ed. Chichester: Wiley, 2010.
[36] O. Tirosh, R. Baker, and J. McGinley, "GaitaBase: Web-based repository system for gait analysis," *Computers in Biology and Medicine*, vol. 40, no. 2, pp. 201–207, 2010.
[37] J. Hamill and K. M. Knutzen, *Biomechanical Basis of Human Movement*. Lippincott Williams & Wilkins, 2006, google-Books-ID: WuWKRc2jZ5AC.
[38] R. Baker, *Measuring walking: a handbook of clinical gait analysis*. London: Mac Keith Press, 2013.
[39] O. Kulyk, R. Kosara, J. Urquiza, and I. Wassink, "Human-centered aspects," in *Human-Centered Visualization Environments*, ser. LNCS, A. Kerren, A. Ebert, and J. Meyer, Eds. Springer, 2007, no. 4417, pp. 13–75.
[40] R. Fabian, "Data-Oriented Design," 2013. [Online]. Available: http://www.dataorienteddesign.com/dodmain/dodmain.html
[41] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of eugenics*, vol. 7, no. 2, pp. 179–188, 1936.
[42] S. G. Eick, "Data visualization sliders," in *Proc. User Interface Software and Technology (UIST)*. ACM, 1994, pp. 119–120.
[43] C. Ware, *Information Visualization: Perception for Design*, 3rd ed. San Francisco, CA: Morgan Kaufmann, 2013.
[44] B. Shneiderman, "The eyes have it: a task by data type taxonomy for information visualizations," in *VL*, 1996, pp. 336–343.

[45] M. Smuc, G. Schreder, E. Mayr, and F. Windhager, “Should we dream the impossible dream of reproducibility in Visual Analytics evaluation?” in *EuroRV3*, W. Aigner, P. Rosenthal, and C. Scheidegger, Eds. EG, 2015.

[46] A. Cooper, R. Reimann, and D. Cronin, *About Face 3: The Essentials of Interaction Design*, 3rd ed. Indianapolis, IN: Wiley, May 2007.

[47] J. Brooke, “SUS–a quick and dirty usability scale,” *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.

[48] J. Nielsen, *Usability engineering*. Boston: Academic Press, 1993.

[49] A. Bangor, P. Kortum, and J. Miller, “Determining what individual SUS scores mean: Adding an adjective rating scale,” *J. Usability Studies*, vol. 4, no. 3, pp. 114–123, 2009.

[50] J. Sauro. (2011) Measuring usability with the system usability scale (SUS): MeasuringU. http://www.measuringu.com/sus.php, [cited: 2016-11-16].

[51] S. Cabral, R. A. Resende, A. C. Clansey, K. J. Deluzio, W. S. Selbie, and A. P. Veloso, “A global gait asymmetry index,” *Journal of applied biomechanics*, vol. 32, no. 2, pp. 171–177, 2016.

[52] A. Muniz and J. Nadal, “Application of principal component analysis in vertical ground reaction force to discriminate normal and abnormal gait,” *Gait & Posture*, vol. 29, no. 1, pp. 31–35, 2009.

[53] C. A. Lozano-Ortiz, A. M. Muniz, and J. Nadal, “Human gait classification after lower limb fracture using artificial neural networks and principal component analysis,” in *EMBC*. IEEE, 2010, pp. 1413–1416.

[54] D. P. Soares, M. P. de Castro, E. A. Mendes, and L. Machado, “Principal component analysis in ground reaction forces and center of pressure gait waveforms of people with transfemoral amputation,” *Prosthetics and orthotics international*, vol. 40, no. 6, pp. 729–738, 2016.

[55] W. Zeng, F. Liu, Q. Wang, Y. Wang, L. Ma, and Y. Zhang, “Parkinson’s disease classification using gait analysis via deterministic learning,” *Neuroscience Letters*, vol. 633, pp. 268–278, Oct. 2016.

[56] J. Christian, J. Kröll, G. Strutzenberger, N. Alexander, M. Ofner, and H. Schwameder, “Computer aided analysis of gait patterns in patients with acute anterior cruciate ligament injury,” *Clinical Biomechanics*, vol. 33, pp. 55–60, 2016.

[57] J. Pauk and K. Minta-Bielecka, “Gait patterns classification based on cluster and bicluster analysis,” *Biocybernetics and Biomedical Engineering*, vol. 36, no. 2, pp. 391–396, 2016.

[58] W. Herzog, B. M. Nigg, L. J. Read, and E. Olsson, “Asymmetries in ground reaction force patterns in normal human gait,” *Medicine and Science in Sports and Exercise*, vol. 21, no. 1, pp. 110–114, 1989.

[59] R. S. Nickerson, “Confirmation bias: A ubiquitous phenomenon in many guises.” *Review of General Psychology*, vol. 2, no. 2, pp. 175–220, 1998.

[60] A. Endert, M. S. Hossain, N. Ramakrishnan, C. North, P. Fiaux, and C. Andrews, “The human is the loop: new directions for visual analytics,” *Journal of Intelligent Information Systems*, vol. 43, no. 3, pp. 411–435, 2014.

Markus Wagner is a research associate at the Institute of Creative\Media/Technologies, St. Pölten University of Applied Sciences (Austria) since 2014. He received his PhD in Computer Science from TU Wien (Austria) in 2017 with highest distinction. His research interests involve knowledge-assisted information visualization and visual analytics methods.

Djordje Slijepčević is a research associate at the Institute of Creative\Media/Technologies, St. Pölten University of Applied Sciences (Austria) and a PhD student at the Faculty of Informatics at TU Wien (Austria). His research focuses on feature learning, pattern recognition and machine learning. He graduated from TU Wien (Austria) with a MSc in Computer Engineering.

Brian Horsak holds a doctorate degree in Biomechanics and Kinesiology. He is researcher at the St. Pölten University of Applied Sciences (Austria) in the Department of Health Sciences. His research focuses on musculoskeletal loading of the lower extremities during locomotion and on innovative technologies supporting patients and therapists during rehabilitation.

Alexander Rind is a research associate at the Institute of Creative\Media/Technologies, St. Pölten University of Applied Sciences (Austria) and a PhD student at the Faculty of Informatics at TU Wien (Austria). Alexander studied Business Informatics in Vienna and Lund and received his MSc degree from TU Wien. His research focuses on interaction and knowledge in visual analytics and visualization for health care.

Matthias Zeppelzauer is a senior researcher at the Institute of Creative\Media/Technologies at St. Pölten University of Applied Sciences (Austria) since 2013. He received his PhD in Computer Science from TU Wien (Austria) in 2011 with highest distinction. His research focuses on multimedia signal processing, pattern recognition and machine learning.

Wolfgang Aigner is the scientific director of the Institute of Creative\Media/Technologies at St. Pölten University of Applied Sciences (Austria) and adjunct professor at TU Wien (Austria). He received his PhD degree and habilitation in computer science from TU Wien in 2006 and 2013. His main research interests include visual analytics and information visualization with a particular focus on time-oriented data.

Question	Answers	Insights
1. Were the filter possibilities understandable?	In general, all participants agreed that the filter possibilities of the system are understandable. P5 stated, "definitely, a restriction of the data makes sense". "You only have to understand the effects of the filters" [P6]. Similarly, P6 argued that she is not sure whether clinicians should or should not use this option.	The integrated filter options for data reduction make sense and are understandable. However, it was asked whether clinicians should use these options.
2. Have the various visualization options contributed to your understanding?	Five out of six participants indicated that the various visualization options contributed to the understanding very well. P1 and P6 stated that the curves and box plots were understandable. P6, however, was confused with the 'Category Difference' because she did not have any previous statistical knowledge in this direction and would therefore need more practice. P5 stated, "it's all pretty clear. It is easy to find out which patients are decisive and it's possible to make restrictions".	The various visualization options, curves and box plots contributed very well and were understandable. Only one participant was confused because of missing statistical background knowledge.
3. Have you understood the use of the knowledge base?	All participants indicated that they have understood the use of the knowledge base, but P2 needed some additional explanation. P3 described that the distribution of the individual parameters of the persons of a category can be viewed based on the matching assignment by an expert and it also allows a great overview of the data. P5 stated that, "it helps to reduce the results and it is very helpful when all have a similar system. You can also assume that a shared database would be good". P4 and P5, it is possible to compare a new data set with the stock in the EKS and then classify it. P3 added, "if I were a clinician in a laboratory, I would add patients".	Generally all participants understood the EKS very well, including the ability to compare new loaded patients to the EKS for classification. Additionally, a single system would be very beneficial as well as a shared EKS.
4. Were the different possibilities of knowledge representation helpful for your decisions?	In general, all participants have classified these options as very helpful. P1 said, "it was very good for quick decision making". "It confirms, or helps, on which one I should concentrate first, so where I set first" [P5]. P3, P4 and P6 described the matching with the data sets as well as the curves as very helpful. P3 added that the twin box plots are very well suited for the overview of the loaded patient to the norm data category and the selected category. This allows the user to click through the data herself, or to rely on the computed matches of the system. The reason for the classification of the loaded patient can also be found on the basis of the knowledge database. P6 said in relation to the twin box plots that she has "no prior knowledge of what deviations I have to look for".	All participants classified the knowledge representation options as very helpful and good for quick decision making. Based on the ITBP it is easy to get an overview of the loaded patient in relation to the norm data category and the selected category. Thus, the analyst has the ability, to klick trough the data herself or to rely on the EKS based matches.
5. How was the expert knowledge represented in the system?	P1, P3, P5 and P6 immediately refer to the categories in the knowledge tree. P3 said, "through the categories, anybody has assigned the patients in groups". P2, P4, and P5 also refereed to the different representations in the knowledge table. The bar charts and graphical summaries show the matching to the individual categories included in the EKS. In addition, P2, P3, P4 and P5 reported in relation to the twin box plots that they can be used to derive the sharp distinction, interval range, variance, differences and relationships for the analysis. P3 added that everything that is not based on the externalized knowledge stored in the EKS is a finding of the analyst.	The participants referred to the categories of the EKS, including patient data of prior analysis, as well as to the different representations in the knowledge table. Additional four out of six participants reported that based on the ITBP the sharp distinction, interval range, variance, differences and relationships can be derived for the analysis.

Question	Answers	Insights
6. Was saving of knowledge based on newly assigned patients or range adjustments of individual values understandable?	All participants have confirmed to have understood the saving process of patients in the EKS. P1 stated that she would not store new patients in categories. She would leave it to other experts, since she did not work on clinical gait analysis on a daily basis. P3 added that it would be a 'nice to have' feature if the category, in which the patient is stored, appears subsequently. P5 stated, "a separate area for patients not yet diagnosed would be helpful." This refers to the saving of patients including them into the matching calculations. Additionally, annotations for the patient data will be helpful for the future.	All participants have understood the saving process to the EKS. The category in which a patient is stored should appear subsequently. As 'nice to haves', a separated saving area for not yet diagnosed patient and the ability for annotations would be helpful. The separated category should not be included in the automated matching calculations.
7. Were the symbols in the knowledge tree structure of the EKS helpful and understandable?	All subjects answered this question with yes. The subjects P2, P5 and P6 initially did not assign any further meaning to the various symbols, but after some practicing in the knowledge tree, they realized that they were different levels of the EKS. P3 stated that the grading symbol size is a good principle. P5 added that the different colors are also very good for distinguishing the levels.	All participants have understood the symbols but most of them did not assign any further meaning to them. Based on the different colors it was easy for the participants to distinguish the different levels of the EKS.
8. Was the graphical summary helpful for the parameters in the category?	Four out of six participants have classified the graphical summary as very helpful. Two participants said that they had difficulties at the very beginning but they were eliminated quickly by a brief explanation. P2 and P6 described that the colors for the distinction are very good, and the individual boxes and their representations contribute to the understanding of the matching results. P2 added that a numbering of the individual parameters in the graphical summary could be helpful in order to retrieve them in the twin box plots.	The graphical summary was classified as very helpful. Only two participants has small issues at the beginning with this visualization metaphor. Based on the coloring and the boxes for the parameters it was easy to understand the matching results. A parameter id related to the ITBPs was added as 'nice to have'.
9. Was the bar chart for the matching helpful?	All participants graded the matching criterion as very helpful, as it provides a very good indication for the analysis starting point. Likewise, it is very helpful to narrow the walking problems of the patient added P5. P6 stated,"it was the first thing I oriented myself".	The matching criteria was indicated as helpful and good starting and orientation point for analysis.
10. Were the shades of the individual values in the range slider helpful and if 'yes', for what?	Five out of six participants immediately recognized the meaning of the shading, and one participant did not pay attention to the shading. P1, P2, P3 and P5 described it as a good basis for the assessment of the range of values (e.g., how the patients are distributed in the class). The darker the shading is the more patients are in this area [P1]. P3 and P5 added that this also gives an overview of the skewness of the distribution in order to recognize deviations. In addition, P5 stated that it can also be seen how well a patient fits into the class or it helps to assess outliers. P6 said that she can explain the shape of the box plots based on the patient distribution.	The range slider shading and its usability for range assessment was quickly recognized by most of the participants. The shading gave an overview of the underlying data derivation and outliers. Additionally, it helps to explain the ITBPs shape.
11. Have you ever worked with box plots before this user study and if 'yes', why?	All participants have confirmed that they know box plots from the statistics in connection with statistical tests or statistical data evaluation.	The participants know box plots from different areas of statistics.

Question	Answers	Insights
12. Were the visualization of the norm data category and the selected category understandable as twin box plot and why?	Five out of six participants have indicated that they have understood the visualization immediately. Additionally, one participant required a brief additional description. P5 said, "yes, is clearly described, contains lots of information in a short time, but its clear." Three out of six participants suggested that the coloring of the box plots while comparing norm data with norm data should change to blue for both contained box plots.	The participants understood the ITBPs immediately. this metaphor contains a lot of information but it is clear structured. In the future, the represented categories should be displayed in their related coloring.
13. Was the comparison of the newly loaded patient, in comparison with the norm data category and the selected category helpful?	Five out of six subjects have indicated that the comparison is helpful. P2, P3 and P4 added that the parameters of the patient can be compared immediately with the parameters of the norm data category and the selected category. It is also possible to see how well the individual parameters of the categories differ [P2, P3 & P4]. This makes it possible to recognize in which category the patient suits best [P4]. P6 has indicated that you need to know how to interpret the presented data.	The comparison of a patient with the norm data category and a selected category based on the ITBPs was described as helpful. It is possible to see the differences of the individual category parameter, but you have to know how to interpret them.

Revision Summary:

KAVAGait: Knowledge-Assisted Visual Analytics for Clinical Gait Analysis

TVCG-2017-02-0028

July 11, 2017

We would like to thank the associate editor and the reviewers of our manuscript for their thoughtful comments and helpful suggestions. These comments helped to improve the quality of the paper even further. We revised the paper accordingly. The most important changes are:

- Case study section revised to clarify the outcomes
- Emphasize benefits and limitations of KAVAGait
- Structural improvements and revisions for readability
- Better illustration of the design rationales
- More clarity on knowledge generation and integration

We put significant effort into this paper revision to address all of the reviewers' suggestions and comments. Please find below a point-by-point response to each comment raised.

Associate Editor (AE)

AE: Less evidence is available in the case study to demonstrate that new insights can be derived by using this developed tool. It seems the case study mainly pertains to validation tasks, which lower the usefulness of the tool.

Thank you very much for this important suggestion. We reformulated and extended the included case study to provide this information now more clearly to the reader (see particularly last paragraph of section 5).

AE: Lack a comparison with the existing tools employed by the experts.

Indeed it is an important issue to relate our new system KAVAGait to currently used solutions. Therefore, we extended the "Problem Characterization and Abstraction" where we provide more detail about currently used tools in this domain. These systems typically present the collected data in a non-interactive interface, as line plots in combination with several calculated discrete parameters. We have also carefully discussed which validation and evaluation methods are suitable in the context of our design study, including a comparative evaluation. However, due to the fact that current systems only provide simplistic visual representations and are lacking further support for clinical decision making, we concluded that a comparative evaluation is not feasible in this case.

AE: Illustrate the design rationale of the relevant visualizations.

The visual representations used in KAVAGait have been developed through continuous refinement in multiple focus group sessions. Typically, we started from sketches based on a small number of known visualization techniques for multivariate time-oriented data. Those suggestions were then discussed in focus group meetings with the domain experts and continuously improved or new ones derived (e.g., ITBP) to fulfil users' needs. We have now added more details about design rationals and have reformulated and reordered several parts in the "Design and Implementation" section. Additionally, we reformulated the "Data-Users-Tasks Analysis" to provide better insights for the readers.

AE: Improve the structure and readability of the paper.

Several efforts were taken to improve the readability as mentioned by the reviewers. We have rearranged the sections "Design and Implementation" and "Reflection and Conclusion". Furthermore, we reduced redundant passages (e.g., in the "Problem Characterization and Abstraction" section) and used more precise language throughout the manuscript.

Reviewer 1 (R1)

R1: I found less evidence from the case study that new medical insights could be created by the system.

Our approach is unique in gait analysis, as this is the first system, which can be used to evaluate gait patterns during clinical practice and store or access "knowledge" simultaneously. The usefulness and the ability to support the process of medical decision making by this accessible knowledge was highlighted as major benefit by a clinical domain expert. However, as we might not have described this clearly enough, we intensively reworked those passages and described the results more clearly.

R1: I found little evidence that unexpected patterns could be discovered, …

As outlined in the comment above, we reformulated and extended the “Case Study” section to provide more details to the reader.

R1: …the authors in 5.2 state that it was a non-goal to compare their design with alternative analysis systems. While the system per se seems useful for comparing and diagnosing, I would have liked to see a description of how the system extends over currently used tools by clinicians and other experts in this kind of data analysis.

This is an important point. Currently available system only provide simple visualization of charts and spread sheets without more sophisticated visualization techniques such as presented in this paper. The concept of storing knowledge and making it available to a clinician during data analysis is a completely new concept to this domain. We described this very early in our manuscript (Introduction, second column, end of first paragraph) and during the “Reflection and Discussion” section in more detail ("Lessons Learned").

R1: I observed that the data set comprises more data than was actually considered in the case study: F_ap and F_ml. Is this data not important, or was there another reason why to show it in the example Figure 2 but not consider it in the case study?

Even though all three components are necessary to describe a patient’s gait pattern in detail, the vertical component of the GRF receives most attention in research and clinical practice. The reason is that the vertical component is the strongest force component and shows greater magnitudes than the other two components (see revised section 3.2). Due to the design study character of our work, we only focused on the vertical GRF in a first attempt. In future work, we will also add these components to the system, besides other data as well. We discuss these future steps and the limitations of the current system in section 6.

R1: Regarding the ITBP diagram, the validation phase indicates this is useful to rank and compare measurements between normal and conditioned patients. Yet, from the whole table of diagrams shown e.g., in Figure 3:3, it seems especially the use of hatching *introduces visual clutter. I was surprised to not find an account of this possible problem in comparing. One may imagine a less cluttered design. However, the authors do not justify their design against alternatives.*

This is a very good point. The design of the HRS is based on the requests of our focus group members to show data in detail. With regards to this, we added a description in the "Problem characterization" section.

R1: ... hence it is not clear that the system can provide generation of novel knowledge to eventually store in the knowledge database (one of the stated goals of the paper).

From a conceptual point of view, facilitating the generation of novel knowledge is certainly a crucial aspect. However, admittedly, the user tests performed so far have focused on confirming intuitions by clinicians (making them explicit and visible) and associating patients to categories. However, as this is a novel system, this kind of adoption behavior is what we have expected. Generation of completely novel knowledge is a subsequent step that entails longer-term usage in the context of a longitudinal study. Moreover, on the technical side, we now have own subsections in the "Design and Implementation" section describing the Explicit Knowledge Store (EKS), the EKS Exploration and Adjustment as well as the Externalized Knowledge Integration and the functionality of the systems knowledge generation loop in detail. We hope this gives clarity to the reader regarding your comment.

R1: As one aspect of exploratory analysis, I would expect that experts may wish to define new measurements to derive from the time series, e.g., to test new medical conditions.

Yes, this is definitely a fascinating avenue to explore gait data. Our group already works on conceptual approaches to derive new data and analysis results, which could be integrated here as well. We therefore, point these possibilities out more clearly in the “Reflection and Discussion” section. This will also be a future task to implement.

R1: Also, searching for patterns may require the user to specify weights for ranking between the different derived measures - is this supported by the system.

The domain experts assured to us that all parameters have the same importance for them for clinical decision-making. Based on this input, we did not include any possibility into the prototype yet to weight these parameters. As we do not know how such rankings would affect the decision making process of a clinician, we were very careful at this point as this could add a strong decision bias by the system.

Reviewer 2 (R2)

R2: The most disputable choice, in my opinion, seemed to be the representation of the 16 values in the EKS and Parameter Explorer. Since these are eight pairs of parameters for the left and right foot, a more harmonious choice for their presentation would have been to have each pair side-by-side, overplayed, or somehow visually comparable. This is done in the Parameter Explorer for categories but not for single patients, which left me wondering about the rationale behind this decision.

Thank you for this important point. We have now extended the "Details on Demand" section, describing the functionality and design rationales of the Interactive Twin Box Plot visualization and the layout. Generally, the comparability of the visualized values is based on the ability to sort them (left stance time / right stance time, left ... / right...). Thus, corresponding parameters for right/left can be displayed directly on top of each other. Based on this sorting, a vertical line is given, supporting the clinicians by comparing the patient's values in relation to left and right. Thus, while it breaks from the left/right metaphor, arranging the parameters vertically allows effective comparison of values on a common scale (cp. Cleveland & McGill, 1984; Heer & Bostock, 2010).

R2: As discussed in 5.3.2, comparing the deviations between the left and right foot of the same patient was proven useful. Hence, could this comparison be done automatically by measuring some kind of patient foot-symmetry score (to be displayed on the overview)?

Yes, this is a good point, gait symmetry and its assessment is crucial for clinicians handling gait data and basing their medical decisions on those. There are several published and well-accepted gait deviation indices, which can be incorporated in our prototype. However, at this stage, these indices have not been implemented yet. We discuss this aspect in more detail in the "Reflection and Discussion" section to be added as an additional supportive parameter for clinical decision making. To support the clinicians we provide a visual representation of gait symmetry between the left and the right foot in our prototype, described in the "Design and Implementation" section in the "Details on Demand" paragraph.

R2: While the manuscript is overall well-written, the readability of the paper can be significantly improved through minor changes.

In line with your comment, we have identified and cleaned up several redundancies in our text and rearranged those paragraphs to enhance the readability. We also made some rearrangements of the sectioning of the paper. In detail, we have eliminated redundancies between the "Problem Characterization and Abstraction" and the "Data-Users-Tasks Analysis" and restructured the sections. Additionally, we reorganized the "Design and Implementation" section to provide a better readability and we added in the "Reflection and Conclusion" a paragraph describing the "Limitations and Future Directions" in relation to our Design Study. Finally, we proofread the complete manuscript with a focus on readability. We are confident that the readability and clarity of the paper strongly increased due to these suggestions and changes.

R2: On the one hand, the paper includes a lot of jargon and abbreviations. Thus, for readers not familiar with the domain-specific problem and the VA approach, this might become an obstacle while going through the paper. While reading the paper, I stumbled upon multiple acronyms, which at first glance seemed unfamiliar, since they were introduced much earlier in the paper. I understand that using them shortens the length of the paper considerably; however, this comes at the cost of a higher mental effort for the reader. As a compromise, I suggest considering to write the unabbreviated words in full-length at the first time they are used in a section as a subtle reminder for the reader to what their abbreviations stand for.

We have now added the full name for each abbreviation once again in the beginning of a main section (as suggested). This will help the readers to keep up to track with the meanings of each acronym.

R2: My major concern regarding the presented approach, however, is the introduced confirmation bias. Since this tool is designed to aid the decision-making of clinicians, it can cautiously be assumed that their judgment is objective due to their expertise. However, the degree of freedom in decision-making provided by the tool allows a subjective steering of the analysis, which can become a self-fulfilling prophecy. Therefore, it is essential for such a tool to provide some visual guidance to the users about the data provenance and certainty. Since this decision is fundamental to the applicability of the tool, I suggest discussing it further in the revised manuscript to prevent misconceptions. ... Did you measure the effects of a confirmation bias?

Currently, our focus lies on the knowledge integration and the interface design to build a supportive tool for clinical gait experts. We share these concerns and agree that avoiding a confirmation bias is an important issue. In the current design, we aim to mitigate a confirmation bias by providing data-driven similarity measures. However, as the category definitions can be adjusted, it is still possible to run into that problem and other measures might be necessary. Moreover, we do currently not measure a possible confirmation bias, but this certainly is a valuable pointer to an avenue of research to be investigated further. We have included a part discussing confirmation bias into the new "Limitations and Future Directions" paragraph in the "Reflection and Discussion" section.

R2: How are the categories determined? Are they theoretically motivated? ... ... Are there theoretical categories that do not have data examples? How are these represented? Does the visualization show the number of participants per category on the overview (to guarantee data provenance)?

All the included categories are based on a sample set provided by the AUVA, the national social insurance for occupational risks for more than 3.3 million employees and 1.4 million pupils and students in Austria. The categories were defined during the "Data-User-Task Analysis" and here especially motivated by the clinical expert from the AUVA, which our team consulted repeatedly during the development. The used categories are part of the categories, which are used during clinical practice in the AUVA. However, of course these are only a sub-set of possible categories. The amount and the specific definition of categories is always strongly related to the patient categories, which an institute regularly sees. In our case, for the AUVA these categories were the most suitable.

R2: *The same applies to parts of the introduction and focus group description. Since all necessary information is included in the paper, this could be resolved by shifting around the order of some sections, e.g., 3.1 and 3.2.*

We have shortened the redundant parts in these sections. Switching the results of the focus group meetings and the "Data-Users-Tasks Analysis" is not possible because the "Data-Users-Tasks Analysis" results are based on the results of the focus group meetings in combination with the performed literature results.

R2: *On the other hand, the second noteworthy aspect for readability improvement is introducing the concepts that are essential for understanding the paper as early as possible in the document. For example, while going through the requirement analysis in section 3.3., the readers are left wondering what a typical analysis workflow of clinicians is. This is only addressed later in the paper but would be helpful for understanding the requirements.*

To support the understanding and the readability of the paper, we now have included a short illustrative example in the "Introduction" section to provide a brief overview of how the clinical workflow looks like.

R2: *How long did the focus group meetings last? In addition, how often did the group meet during the 13-month time period (that is mentioned in the paper)?*

We reformulated this part to present the detailed amount and timespans for our performed meetings.

R2: *How was Equation 1 derived? Are there other reasonable alternatives to measure the matches, which might also make sense to the domain experts?*

We have included a more detailed description for Equation 1 in the "Visualization Concept" section.

R2: *Why do clinicians have to adjust the ranges of the HRS? Are the filtering and adjustments kept for more than one analysis session?*

We have extended the description "EKS Exploration and Adjustment". If a category is empty or contains not enough data sets for example, the clinician has the ability to integrate his / her implicit knowledge directly into this category to create a value range for the automated matching. The settings that will be done in relation to the categories in this area are stored as explicit knowledge for all future sessions. However, it is also possible to delete them in the future and recalculate the ranges based on the included patients.

R2: Why does the EKS not include prototypical (aggregated) GRF plots for every category in addition to the 16 parameters?

This is a very interesting suggestion that we also considered. However, in the context of our problem analysis, our clinical expert collaborators from AUVA expressed that it is very important for them to see and work with the individual (raw) data.

R2: Does every patient strictly belong to one category (after being confirmed by an expert) or can a patient be added to multiple categories?

A patient can also belong to more than one category, depending on his or her gait abnormalities. Therefore, the system provides the ability to add a patient to more than one category. In this case, it is important to maintain the individual characteristics of each category based on the included parameters. To avoid the overlapping of categories, a clinician also has the ability to include only relevant parameters of the patient to a specific category.

R2: In addition to matching categories, can clinicians query for the most similar patients to a selected person? This would probably be useful to compare the effectiveness of past treatments on particular cases.

This is a very good idea, which we were discussing as well as a future direction for our proposed system. We added a short paragraph discussing this in the last section "Reflection and Discussion".

R2: Other minor issues are:

- *'Loaded person/patient' → selected patient (?)*
- *Caption Fig.4: 'a specific gait abnormalities' → 'a specific gait abnormality'*
- *page 8, line 30, left column: 'vice versa' → 'respectively'*
- *page 8, line 57, left column: 'Thereby' → 'Hereby'*
- *Caption Fig.5: 'including the patient' → 'that includes the patient'*
- *page 11, line 41, left column: 'as very as very'*
- *page 11, line 27, right column: 'selection of the 'Norm Data Category' for comparison in the ITBP Section 4 in the future.' → sentence doesn't make sense to me*

Thank you for pointing out these issues – they have all been corrected accordingly.

Reviewer 3 (R3)

R3: It is a good choice to refer to the nested model presented by Munzner. This paper explained how to complete the design process in accordance with the nested model in detail. However, the narrative is a little chaotic. Section 3.1.2 discusses the many results you get from the focus group. I think the discussion can be generalized here, and then, the specific information can be described in detail based on the model of 'Data-Users-Tasks Analysis' in Section 3.2. Actually, those two parts have a little overlap. Moreover, Section 3.1.3 can be integrated in the introduction of 'tasks'. In such a way, the readers will have a better understanding of the nested model you employed.

We have eliminated the redundant elements in these sections. Switching the results of the focus group meetings and the "Data-Users-Tasks Analysis" is not possible because the "Data-Users-Tasks Analysis" results are based on the results of the focus group meetings in combination with the performed literature results. However, we made several changes to the text to increase the readability of the paper.

R3: In Section 1, the contribution 1-3 do not sound so appealing individually. All of them are common to visual analytics systems. To highlight the design process, the contribution 1-3 can be re-organized to make the work sounds more solid.

We have extended the four contributions of the paper. In general, our intent is to structure the paper's contribution along Sedlmair et al. (2012) "Design Study Methodology" to show the importance of a problem characterization and abstraction, the validated design of a prototype and its reflection, which are all important aspects for visual analytics systems.

R3: The selected reference is not very close to this paper. Compared with many studies related to crowd trajectory analysis (like [11] in the paper) I think [1] is more corresponding to this work. [1] Jang S, Elmqvist N, Ramani K. MotionFlow: Visual Abstraction and Aggregation of Sequential Patterns in Human Motion Tracking Data [J]. IEEE transactions on visualization and computer graphics, 2016, 22(1): 21-30.

Thank you for the suggestion of the relevant reference. It was incorporated in the Related Works and in the References as *[17] S. Jang, N. Elmqvist, and K. Ramani, "Motionflow: Visual abstraction and aggregation of sequential patterns in human motion tracking data," TVCG, vol. 22, no. 1, pp. 21–30, Jan 2016.*

R3: In addition to the nested model, this paper also refers to another classic framework: the interaction pipeline provided by Shneiderman. All descriptions about the interactions involves the same figure. I wonder if Figure 3 can include all interactive effects. If not, it is better to provide more figures to illustrate.

We have added additional references to several figures to show all the functionalities of our prototype. Additionally, we are providing an introduction video for KAVAGait on our supplement material website: https://phaidra.fhstp.ac.at/detail_object/o:1928.

R3: Besides, as the most important innovation in the visualization part, the Interactive Twin Box Plot is neither easy to understand nor beautiful as shown in Figure 4. The three parts of ITBP should have margins between each other to provide a clear information.

We carefully considered the design of ITBP: the three parts share a coordinate system and are to be used together. The parts are clearly separated by colors that have an associated meaning in context of the application (blue := norm, orange := pathological condition). An alternative design with margins between its parts would perceptually break the component apart (tested in preliminary studies). Since the ITBP is used within a table, i.e. in direct vicinity of other ITBPs, it would be hard to recognize which parts belong together.

R3: Also, note that there is text overlap in Figure 4.

We refined our source code, which was responsible for this overlap and recreated this figure.

R3: Note that there are many figures cannot show information clearly because of the small space (Figure 5, 6, 7). The related Figure needs to be re-designed for effective space using.

To represent the included figures more clearly, we have increased the space usage of Figure 5 and Figure 6. Figure 7 is a representation of the included knowledge loop whereby the small images of the system should be seen as thumbnails. More material (in full resolution) can be found on our supplement material website: https://phaidra.fhstp.ac.at/detail_object/o:1928.

R3: In Section 5.1.1, you introduce three experts with a table and label them as E1, E2 and E3, however, such labels are not mentioned latter. I think you had better distinguish the evaluation of each expert or omit the labels.

We have removed the table from the paper in order to create space for more important content.

R3: Subsequently, when reading the results of the experiment, I am concerned about the results of the interview. In addition, I expect to see quantitative statistical results. However, there are few related introduction and discussion.

A number of evaluation methods are commonly used in visualization/visual analytics and each evaluation method has its own specific limitations. Since KAVAGait is a solution for expert users, we focused our evaluation on domain experts and were successful to recruit six domain experts for the user study and one leading national expert for the case study. Under these circumstances, semi-structured interviews with qualitative analysis yield richer results than a quantitative experiment would have (for which the sample size would be too small). For this reason, we decided for a qualitative evaluation. More detailed results of the interviews can be found in the supplemental material.

R3: One of another concern is the novelty of this paper. As it follows a nested model for design study, I would expect to see some discussions, comparisons and analysis of related works, and other designs (if there are), and failure cases.

We expanded the discussion of design alternatives and comparisons to related work (esp. in regards to knowledge assisted visual analytics). An even more elaborated discussion would exceed the page limit of the journal. However, we are confident that all important points are now adequately addressed and that the changes made in line with the reviewers' comments helped to significantly increase the quality of our work.

We once again would like to thank the associate editor and the reviewers for their thoughtful comments and suggestions that helped us to enhance the quality of our submission.

Best regards,
The authors