Laban Movement Analysis using Kinect

Bernstein Ran[1], Shafir Tal[2], Tsachor Rachelle[3], Studd Karen[4], and Schuster Assaf[1]

[1]Department of Computer Science, Technion I.I.T, Haifa, Israel
[2]The Graduate School of Creative Arts Therapies, University of Haifa
[3]School of Theatre & Music, The University of Illinois at Chicago
[4]School of Dance, George Mason University

Abstract—Laban Movement Analysis (LMA), developed in the dance community over the past seventy years, is an effective method for observing, describing, notating, and interpreting human movement to enhance communication and expression in everyday and professional life. Many applications that use motion capture data might be significantly leveraged if the Laban qualities will be recognized automatically. This paper presents an automated recognition method of Laban qualities from motion capture skeletal recordings and it is demonstrated on the output of Microsoft's Kinect V2 sensor.

Keywords—Laban Movement Analysis, Kinect, Machine Learning.

I. INTRODUCTION

A. Laban movement analysis

LMA is a formal language for motion description first developed by Rudolf Laban [13] and colleagues in the middle of the 20th century. LMA describes both conscious and unconscious human movement, based on Laban's categories of *Body*, *Effort*, *Shape*, and *Space*. LMA has been used in the fields of dance, acting, athletics, physical therapy, and psychology and behavioral science. LMA helps actors create momentary moods and portray personality traits through movement. For example, LMA work investigates the *Effort* properties *Flow*, *Space*, *Time* and *Weight* of all movement and helps actors think specifically about why their character might move in a jerky, fast, light and direct manner versus a heavy, slow, indirect and uninterrupted manner. The entire LMA hierarchy is shown in figure 1.

B. Motivation for Automated LMA

There are numerous applications for computerized identification of the qualities that characterize each possible human movement. Examples include the generation and control of specific expressive movements of avatars, virtual characters, or robots in mixed reality scenarios [14]; detection of personality traits during a job interview [17]; early detection, severity assessment or revealing of genetic tendency (phenotype) towards various illnesses such as Alzheimer's, autism, Parkinson's disease [2], or schizophrenia, based on analysis of the person's motor behavior. Automated emotion recognition from movement is another important application, which may have a variety of uses such as online feedback to presenters to help them convey through their body language the emotional message they want to communicate (e.g., politicians and public speakers or actors in training) [5]; or recognition of people's emotions during interactive games such as those played using the Xbox [21].

For reducing our data collection and analysis effort, we focused our work on 18 Laban qualities (as listed in Diagram 6) that have been found predictive for emotional state [23].

C. Related Work

Several attempts were made to recognize Laban qualities. The first was Chi et al. [1], who quantified *Effort* and *Shape* for animation. Most of the other attempts were for emotion recognition in the context of Human Robot Interaction (HRI). Martin et al. [8] analyzed the importance of gestures in emotion recognition for HRI. Masuda et al. generated emotional body motion for a human form robot [14]. Rett et al. proposed a human motion recognition system using a Bayesian reasoning framework [20]. The second line of works focused on LMA (not on emotions), but not using Kinect. Lourens et al. [3] used video data and Samadani et al. [4] used a high quality MOCAP camera, but both of them analyzed only hand gestures. A third line of works used Kinect as the main sensor for skeletal information. Gabel et al. [11] used Kinect for gait analysis. The work of Zacharatos et al. [21] was inspired by LMA for emotion recognition using Kinect. His feature extraction method was influenced by LMA principles, but he did not attempt to recognize the qualities themselves. Kim et al. [16] did attempt to
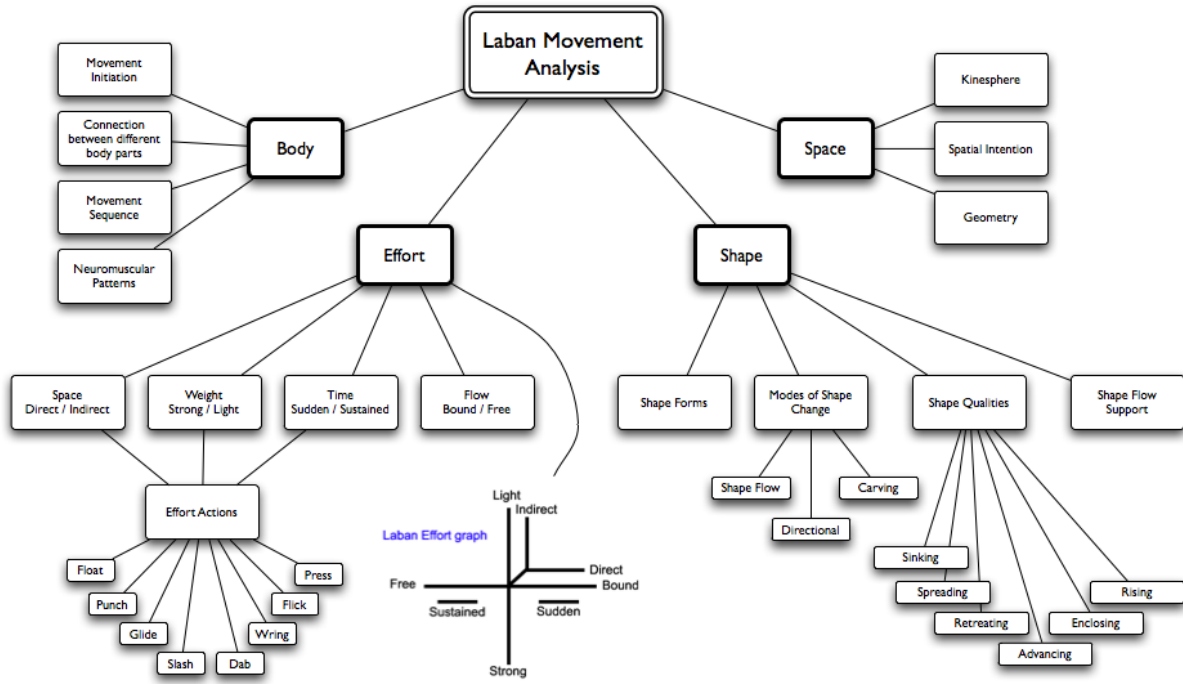
Fig. 1. Main axes of LMA. Taken from [22]

II. METHOD

do so but not on a real dataset and their work did not include a performance evaluation.

Because we are the first to handle Laban recognition with Kinect, we had to create a dataset from scratch. To reduce the noise, and ensure that we capture the essence of the Laban qualities in our dataset, we decided that most of it should be built by recording several Certified [Laban] Movement Analysts (CMA), with just a few validation clips taken from recordings of ordinary people. We did not want to constrain the lengths of the clips to be equal, so in order to get feature vectors of uniform length (regardless of the original length of the clips), every feature is function of a whole clip (for example, the variability of the elbow's acceleration). On the uniform length feature vector we applied feature selection, single task learning (learning a model for every quality separately), and multitask learning (learning a model for all the qualities together).

A. Clip Collection

Two main datasets were collected:

- CMA dataset - includes 6 CMAs performing in about 80 clips each (a total of 550 clips). Every clip is about 3 seconds long, and the CMAs executed combinations of the 18 qualities. To achieve uniform distribution of the Laban qualities over the dataset, in every clip the CMA was asked to perform actions that include several specific qualities, and nothing but them.
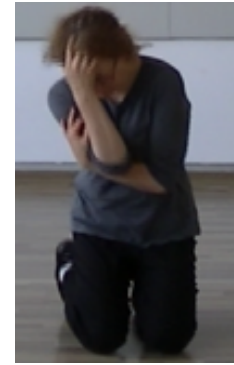

Fig. 2. CMA during a clip

- Non-CMA dataset - includes 2 subjects without a background in movement analysis, performing 30 clips each. Every clip is also about 3 seconds long, and the subject was asked to perform one out of several tasks.

B. Clip Labeling

To achieve a ground truth labeling for the two datasets, every clip was tagged by a committee of 2 CMAs who determined which Laban qualities appear in the clip. The use of a committee decision instead of the subjective opinion of one CMA decreases the labeling noise and the decision is considered as ground truth.

C. Feature Extraction

Due to unequal length of clips, all the features that were extracted are in whole clip granularity.

1) Primitive Features

For every joint in the skeleton the angular velocity, acceleration and jerk were extracted, and for each one of them the mean, variance, skew and kurtosis were extracted (the extraction of the last four moments is denoted as ϕ).

$We\ denote_j(t)$ as the vector (as we get it from the Kinect) of the position of joint j in time t in a clip with n frames, and α_j is a coefficient proportional to the mass around the joint.

2) Shape Analysis: Sagittal Plane

Laban shape analysis of the sagittal plane is based on the distinction between two qualities, *Advance* and *Retreat*. This distinction was quantified by projecting the velocity vector of the Center of Mass (CM) on the vector of the front of the body. The CM was approximated in this case by the average of all the joints. The front of the body was approximated by the perpendicular vector to the vector between the Left Shoulder (LS) and the Right Shoulder (RS).

If sag stands for sagittal, then from the definition of CM of a physical system,

$$\vec{P}_{CM}(t) = \sum_{j \in Joints} \alpha_j \vec{P}_j(t),$$

$$\vec{P}_{shoulders}(t) = \vec{P}_{LS}(t) - \vec{P}_{RS}(t),$$

the front is perpendicular to $\vec{P}_{shoulders}$, so we can easily calculate it with:

$$\vec{P}_{front} = \vec{P}_{shoulders} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix},$$

$$S_{sag}(t) = \vec{P}_{CM}(t) \cdot \vec{P}_{front}(t),$$

$$\vec{F}_{sag} = \phi([S_{sag}(1), \dots S_{sag}(n)]),$$

where ϕ was denoted in the beginning of the section.

3) Shape Analysis: Horizontal Axis

Here the distinction is between *Spreading* and *Enclosing* on the horizontal axis. This distinction was quantified by measuring the average distance between every joint to the vertical axis of the body that extends from the Head (H) to the Spine Base (SB).

$$d_j = \frac{\left| (\vec{P}_j - \vec{P}_{SB}) \times (\vec{P}_j - \vec{P}_H) \right|}{\left| \vec{P}_H - \vec{P}_{SB} \right|},$$

$$S_{horiz}(t) = \sum_{j \in Joints} d_j(t),$$

$$\vec{F}_{horiz} = \phi([S_{horiz}(1), \dots S_{horiz}(n)]),$$

4) Shape Analysis: Vertical Axis

Here the distinction is between *Rise* and *Sink* on the vertical axis. This distinction was quantified by measuring the average distance on axis y of each joint from the CM. This quantification is based on the assumption that the body is "longer" when rising.

$$S_{vert}(t) = \sum_{j \in Joints} \left| \vec{P}_j - \vec{P}_{CM} \right|,$$

$$\vec{F}_{vert} = \phi([S_{vert}(1), \dots S_{vert}(n)]),$$

5) LMA Effort Analysis: Time Category

Here the distinction is between *Sudden* and *Sustained*. This quality was quantified by the skew of the acceleration, relying on the assumption that the acceleration of a sudden movement will be skewed further to the left, i.e., will get a higher value at the beginning of the movement.

6) Effort Analysis: Space Category

Here the distinction is between *direct* and *Indirect* motion. This quality was quantified by the angle between the movement vector of a joint to the next one, relying on the assumption that in direct movement every vector will be in the same direction as the last (the angle between them is small). The velocity direction V is calculated by $\vec{V}_j(t) = \vec{P}_j(t+1) - \vec{P}_j(t)$, and the angles between a direction to the next one is calculated with the inner product $\vec{A}_j(t) = \vec{V}_j(t+1) \cdot \vec{V}_j(t)$.

D. Performance Evaluation

From a statistical point of view, we have 18 possible labels (Laban qualities) for every clip. Each clip was a combination of just a few of these, often 3-4, which means that there is about an 85% chance that a quality won't appear in a clip. Due to this sparsity, accuracy alone is not a relevant metric for the performance

evaluation because one can get 85% accuracy by stating that for every recording none of the qualities appear. A better evaluation would have to combine the precision and recall rates of the classifier. This can be done using the F1 score:

$$F_1 = \frac{2 \cdot precision \cdot recall}{precision + recall}.$$

E. Feature Selection

Every clip is extracted into a vector of 6120 features, most of which are noisy or redundant, thus requiring massive feature selection. The feature selection is done in three stages:

- Computing the Anova F-value for every feature over the training set. Cross-validation was used to determine the optimal number of features that should be left. As seen in Figure 3, filtering out most of the features yielded better results than not filtering them, where using the top 4% of features was optimal.
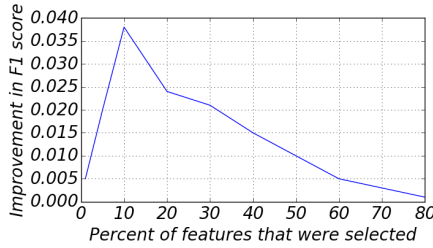
Fig. 3. Influence of the number of features on the performance. The selection was made according to statistical significance. The blue line is the difference between the score with and without feature selection. It can be seen that the optimal fraction of features to select is 4%.

- The second phase of feature selection was conducted by Information Gain (IG) rating of the features. As seen in Figure 4, the optimal ratio was obtained by selecting the top 60% out of the features that remained after the first phase of feature selection. Examples of qualities and their most significant feature are given in Table I. The "Information Gain" metric used in the table is defined as:

$$IG(T, f) = H(T) - H(T|f),$$

  where T is the training set, f is a feature, and H() is the information entropy of a dataset.
- The third phase of feature section was conducted using the Least Absolute Shrinkage and Selection Operator (LASSO) regularization.
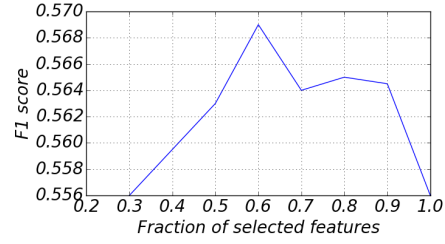
Fig. 4. Influence of the number of features selected with IG from the subset of features chosen in the first phase on the performance. The optimal ratio was 60%.

III. EXPERIMENTAL SETUPS AND RESULTS

A. Multilabel Classification

Multilabel learning deals with the problem where each instance is associated with multiple labels simultaneously, where the number of labels is not fixed from instance to instance. The task of this learning paradigm is to predict the label (Laban quality) set for each unseen instance (skeletal recording), by analyzing training instances with known label sets. The multilabel approach taken in this paper is to break the LMA problem into 18 binary classification tasks — one for every Laban quality — where every binary decision is whether or not the quality exists.
The following subsections will describe several experimental setups where the results in each will serve as a baseline for the next.

B. Per CMA Evaluation

In this experiment the train and test datasets are taken from the same CMA. The performance on every Laban quality separately is demonstrated on a dataset of one of the CMAs in Figure 6. In Figure 5 the incremental evolution of the algorithm is described from step to step with the next notation:

- *Chance* stands for randomly tossing a balanced coin in every classification decision.
- *NN* stands for applying the Nearest Neighbors algorithm.
- *LinearSVC* stands for Support Vector Classifier (SVC) with a linear kernel.
- *LabelBalancing* stands for giving greater weight to clips that contain the quality due to the small fraction of them in the whole dataset.
- *Lasso*, *SFS* (Statistical Feature Selection), and *InfoGain* (information gain based feature selection) were described in the *Feature Selection* section.

IV. CONCLUSION

We developed a method for recognizing Laban qualities using the Microsoft Kinect sensor. We developed

Quality	Feature Description	Information Gain	p-value
Jump	Vertical relative position of center of mass	0.28	3.00E-50
Spread	Horizontal relative position of left elbow	0.25	2.20E-49
Rotation	Left shoulder horizontal acceleration	0.21	4.80E-46
Up+Rise	Vertical relative position of left elbow	0.2	3.00E-31
Free+Light	Left elbow angle's variability	0.16	1.10E-24
Rhythmicity	Vertical relative position of center of mass	0.16	1.50E-11

TABLE I

EXAMPLE OF SEVERAL QUALITIES AND THE FEATURE FOUND TO BE THE MOST INFORMATIVE FOR THEM. "RELATIVE POSITION" STANDS FOR THE POSITION OF THE JOINT RELATIVE TO THE ANCESTOR JOINT IN THE JOINT HIERARCHY.
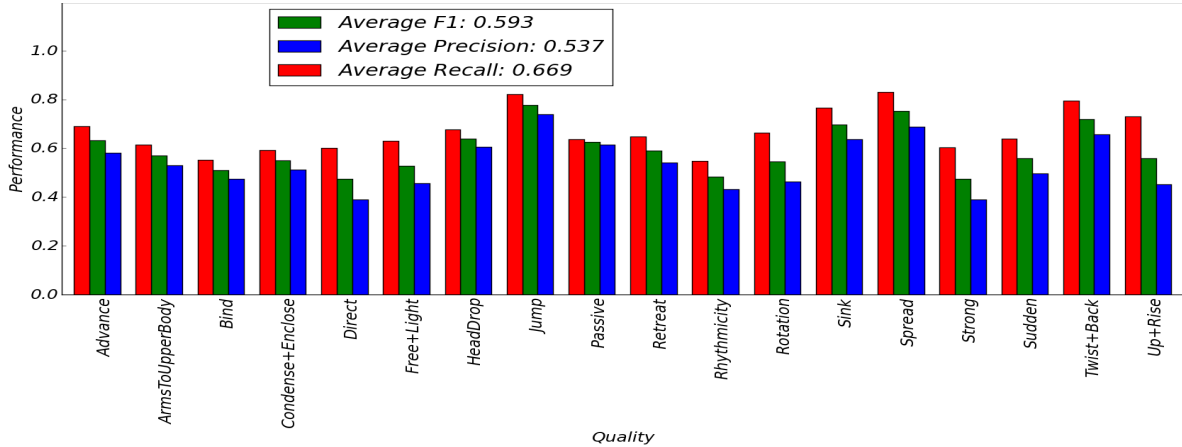

Fig. 6. Recall, precision and F1 score of each Laban quality separately. The evaluation was conducted on a dataset that was captured on only one CMA.
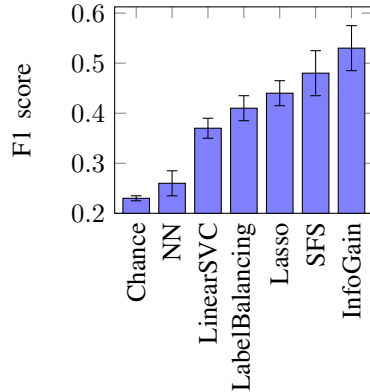

Fig. 5. Evaluation of every CMA's dataset separately in the single task learning setting. Each column represents an additional step in the algorithm's evolution. The results are the average F1 score and its standard deviation (STD) between the CMAs.

a method for recognizing Laban qualities using the Microsoft Kinect sensor. Our method obtained a recall and precision of about 60% over the qualities. The larger movements, such as *jump*, *spread*, and *sink*, are easier to quantify, and hence easier to recognize (precision and recall of 60-90%). The more subtle qualities, such as *strong* and *passive*, are harder for us to quantify in kinematic measurements, which causes a degradation in the performance (precision and recall of 40-60%). Overall we believe that we succeeded in capturing the essence of most of the qualities, using a cheap ($100) and widely available sensor. We believe that our work will provide the foundation and inspiration that will make the LMA method applicable in many more methodologies and processes.

REFERENCES

[1] Diane Chi, Monica Costa, Liwei Zhao, and Norman Badler. The emote model for effort and shape. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 173–182. ACM Press/Addison-Wesley Publishing Co., 2000.

[2] Antonio Camurri, Barbara Mazzarino, Gualtiero Volpe, Pietro Morasso, Federica Priano, and Cristina Re. Application of multimedia techniques in the physical rehabilitation of parkinson's patients. *The Journal of Visualization and Computer Animation*, 14(5):269–278, 2003.

[3] Tino Lourens, Roos Van Berkel, and Emilia Barakova. Communicating emotions and mental states to robots in a real time parallel framework using Laban movement analysis. *Robotics and Autonomous Systems*, 58(12):1256–1265, 2010.

[4] Ali-Akbar Samadani, Sarahjane Burton, Rob Gorbet, and Dana Kulic. Laban effort and shape analysis of affective hand and arm movements. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 343–348. IEEE, 2013.

[5] Anh-Tuan Nguyen, Wei Chen, and Matthias Rauterberg. On-

line feedback system for public speakers. In *IEEE Symp. e-Learning, e-Management and e-Services*, 2012.

[6] Afra Foroud and Sergio M Pellis. The development of roughness in the play fighting of rats: A Laban movement analysis perspective. *Developmental Psychobiology*, 42(1):35–43, 2003.

[7] Fabian Pedregosa and Gael et al. Varoquaux. Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research*, 12:2825–2830, 2011.

[8] Jean-Claude Martin, Radoslaw Niewiadomski, Laurence Devillers, Stephanie Buisine, and Catherine Pelachaud. Multimodal complex emotions: Gesture expressivity and blended facial expressions. *International Journal of Humanoid Robotics*, 3(03):269–291, 2006.

[9] Yoren Gaffary, Jean-Claude Martin, and Mehdi Ammi. Perception of congruent facial and haptic expressions of emotions. In *Proceedings of the ACM Symposium on Applied Perception*, pages 135–135. ACM, 2014.

[10] Yoren Gaffary, Victoria Eyharabide, Mohamed Yacine Tsalamlal, Jean-Claude Martin, and Mehdi Ammi. Comparison of statistical methods for the analysis of affective haptic expressions. In *Haptic and Audio Interaction Design*, pages 69–78. Springer, 2013.

[11] Moshe Gabel, Ran Gilad-Bachrach, Erin Renshaw, and Assaf Schuster. Full body gait analysis with kinect. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 1964–1967. IEEE, 2012.

[12] Rachelle Tsachor. Laban/Bartenieff based somatic movement therapy: One practitioner's methodology and individual case studies. *In The Journal of Laban Movement Studies*, 3(1):20–53, 2013.

[13] Rudolf Laban and Lisa Ullmann. The Mastery of Movement. 1971.

[14] Megumi Masuda and Shohei Kato. Motion rendering system for emotion expression of human form robots based on Laban movement analysis. In *RO-MAN, 2010*, pages 324–329. IEEE, 2010.

[15] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.

[16] Woo Hyun Kim, Jeong Woo Park, Won Hyong Lee, Hui Sung Lee, and Myung Jin Chung. Lma based emotional motion representation using rgb-d camera. In *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction*, pages 163–164. IEEE Press, 2013.

[17] Jacqyln A Levy and Marshall P Duke. The use of laban movement analysis in the study of personality, emotional state and movement style: An exploratory investigation of the veridicality of" body language". *Individual Differences Research*, 1(1), 2003.

[18] Tal Shafir, Stephan F Taylor, Anthony P Atkinson, Scott A Langenecker, and Jon-Kar Zubieta. Emotion regulation through execution, observation, and imagery of emotional movements. *Brain and cognition*, 82(2):219–227, 2013.

[19] Tal Shafir and Rachelle Tsachor. Something in the way she moves... unique sets of movement characterists are associated with and enhance specific emotions. poster accepted for the international convention of psychological science (ICPS), 2015.

[20] Jörg Rett and Jorge Dias. Human-robot interface with anticipatory characteristics based on Laban movement analysis and bayesian models. In *Rehabilitation Robotics. ICORR, IEEE 10th International Conference on*, pages 257–268. IEEE, 2007.

[21] Haris Zacharatos, Christos Gatzoulis, Yiorgos Chrysanthou, and Andreas Aristidou. Emotion recognition for exergames using Laban movement analysis. In *Proceedings of the Motion on Games*, pages 39–44. ACM, 2013.

[22] http://www.laban-movement-analyses.be.

[23] Private consultation with Shafir Tal and Tsachor Rachelle.