Towards the use of consumer-grade electromyographic armbands for interactive, artistic robotics performances

Ulysse Côté-Allard[1], David St-Onge[2], Philippe Giguère[3], François Laviolette[4] and Benoit Gosselin[5]

***Abstract*— In recent years, gesture-based interfaces have been explored in order to control robots in non-traditional ways. These require the use of systems that are able to track human body movements in 3D space. Deploying Mo-cap or camera systems to perform this tracking tend to be costly, intrusive, or require a clear line of sight, making them ill-adapted for artistic performances. In this paper, we explore the use of consumer-grade armbands (*Myo armband*) which capture orientation information (via an inertial measurement unit) and muscle activity (via electromyography) to ultimately guide a robotic device during live performances. To compensate for the drop in information quality, our approach rely heavily on machine learning and leverage the multimodality of the sensors. In order to speed-up classification, dimensionality reduction was performed automatically via a method based on Random Forests (RF). Online classification results achieved 88% accuracy over nine movements created by a dancer during a live performance, demonstrating the viability of our approach. The nine movements are then grouped into three semantically-meaningful *moods* by the dancer for the purpose of an artistic performance achieving 94% accuracy in real-time. We believe that our technique opens the door to aesthetically-pleasing sequences of body motions as gestural interface, instead of traditional static arm poses.**

I. INTRODUCTION

As robots are increasingly present in daily life, both in the industrial and private spheres - their means of interaction with humans is of growing significance. Remote control devices (joysticks, computer-based control application, smart phones, etc.) require a user to consciously transcript their commands into specific actions on the apparatus. For many contexts, this may be a distracting constraint or require too much time to learn for a user. This work was conducted for a specific category of users: artists. A wealth of performers have been engaging in robotic performances for more than fifty years now [1], [2]. Artists have always been explorers of human feelings and perception [3]. As for many media or digital arts performances, audiences gathering around such pieces are usually aware of what they are going to see, which mean that they are more prepared to question the work and think on it than general audiences. In this context, the artist, for instance a dancer, cannot effectively focus on delivering a performance while issuing non-intuitive commands to a robotic device. Furthermore, the guidance of the robot has to be done in an inconspicuous manner to the spectator, so as not to become a technical demonstration. Consequently, more intuitive devices that could be integrated seamlessly into an artistic performance have to be developed. For instance, the pitch of the human voice has been employed to control the motion of cubic blimps [4], fitting the aesthetic context of a singing/dancing performance (Fig. 1).

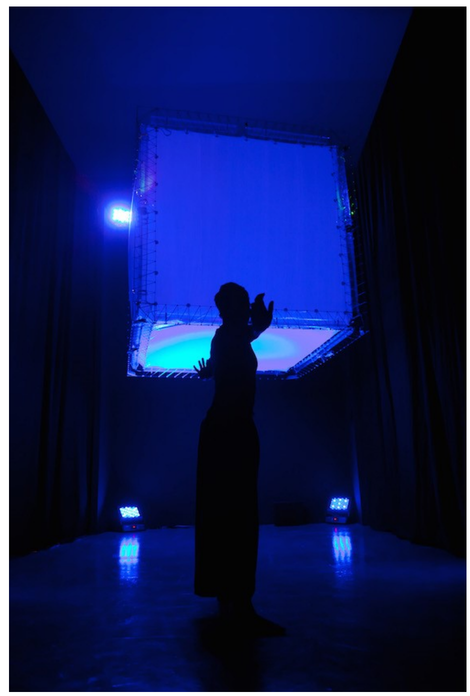

Fig. 1. Dancer Ghyslaine Doté with a cubic blimps flying robot in a performance in São Paulo, 2012. Employing large robotic blimps during artistic performances severely reduce the applicability of line-of-sight sensors such as Mo-cap and cameras.

The last decade of research has led to a number of original ways to interpret human behavior such as visual movement detection, speech recognition, haptic devices (force sensing), smart watches, etc [5]. Interfacing these with robots, in order to intuitively control them, remains a challenge that is being addressed in the active field of Natural User Interface [6], [7]. However, even Natural User Interfaces often constrain the creativity when technically too limited for the artist intent [8]. Indeed, extracting pre-processed interaction actions from features of any data is less flexible to the artist uniqueness and his fatigue throughout a performance [9]. Thus this paper

[1]Ulysse Côté-Allard with Faculty of Electrical Engineering, Laval University, Qc, Canada ulysse.cote-allard.1@ulaval.ca

[2]David St-Onge is with Faculty of Mechanical Engineering, Laval University, Qc, Canada david.st-onge.2@ulaval.ca

[3]Philippe Giguère is with Faculty of Computer Science and Software Engineering, Laval University, Qc, Canada philippe.giguere@ift.ulaval.ca

[4]François Laviolette is with Faculty of Computer Science and Software Engineering, Laval University, Qc, Canada francois.laviolette@ift.ulaval.ca

[5]Benoit Gosselin is with Faculty of Electrical Engineering, Laval University, Qc, Canada benoit.gosselin@gel.ulaval.ca

presents a new take on employing human body movements to enable performers to seamlessly control robots the projected emotions of their body. To that effect, experiments including two professional dancers were conducted with the goal of assessing the accuracy of an adaptive lightweight system for dance recognition.

In gesture recognition, the focus is on accurately detecting a number of simple gestures within a small, pre-determined workspace. For example, the Leap Motion Controller™ can track and recognize hand and finger gestures directly in front of a computer screen [10]. The well-known Kinect™ systems can perform hand and arm gesture recognition [11] with some limited finger movement tracking. It has successfully been used to command motion primitives on quadrotors [7]. For wider workspaces and movements, a motion capture system such as the Vicon™ is better suited. It is however expensive, and often cumbersome, to set up unsightly cameras in the context of performing arts. Some research groups have tackled the problem of whole body recognition in larger space through the use of stereo-camera rigs [12] or monocular camera systems for low resolution gestures at long range [13]. All of these systems pose the problem of necessitating an unobstructed line of sight, which can be hard to obtain, notably in street performance settings.

On the other hand, EMGs give information on the wearer's motions and gestures from a more intimate, proprioceptive perception instead of an absolute external recognition. Furthermore, it removes the requirement of clear line of sight. However, these signals are challenging to interpret.

Attempts in the art domain to employ these devices have already been made. Artists are questioning and exploring every possible mean to interface their body with our world so as to render and create emotions. The German group Palindrome utilized EMG for visual amplification of a dancer's movements [14]. Bill Vorn from Concordia University in Canada created the *Grace State Machine*, a parallel platform mechanism stacked in columns moving according to the analog biometrics of a dancer, among which were EMG sensors.

EMGs are well known in medical applications for diagnostics, and also interfacing with prosthetic devices and wheelchairs. For the latter, neural and Bayesian networks have been successfully applied to classify hand gestures [15], [16]. For a neural network, a solution to the time-varying factor of the EMGs was proposed [17]. However, these algorithms require very large datasets and powerful processing for training. For a live performance with choreographers, the classifier must be trained within minutes on a standard computer to allow for rapid experimentation and exploration of *ad-hoc* gestural interfaces. Faster algorithm such as Random Forest were also explored, but only in the context of precise grasping tasks [18]. The unique context (dance and choreography) of this study is explained in the next section.

Because no currently available system meet the requirements for a live adaptive mapping of body movement to control a mechanic device, this paper introduce a fully tested solution. Its main contribution is to provide the user with a fast live training machine, lightweight, making the interpretation of dancers moods easy and flexible.

II. CONTEXT OF THE PROJECT: DANCE AND CHOREOGRAPHY

Human body language is complex and meaningful. In robotic applications, the focus is most often on simple gesture recognition [19]. However, whole body motion can bring significantly more information, such as intention and mood. Dance is an active and conscious activity playing on the potential of human body motion. Dancers and choreographers are trained by profession to identify the connections between body movements and its mapping to the audience's emotions. Without this automatic understanding and innate emotional encyclopedia, this ability is impossible to learn through a methodical approach. Thus, the designation of body movements and their emotional association should primarily be driven by the dancers and choreographers. With such knowledge, it is possible to design a system reacting to *perceived* emotions by categorizing the body movements.

Central to this work, the concept of a dance mood is taken from the choreography discipline. Choreographic theory identifies four key elements in dance: design, dynamics, rhythm and form [20]. Dances can be analyzed and designed in terms of these basic elements to generate moods. Thus moods are clusters of dance movements, arbitrarily defined by a dancer or a choreographer, based on the perceived emotion that they each create. The final hypothesis of our classifier is that it can be trained on those groups. Since moods are specific to a given state of the dancer (psychological state, social context, creation environment), it is complex to obtain large number of samples and thus limit the potential training algorithms fit for moods recognition.

To capture the dancer's movement in a non-obtrusive manner, the device must be small, lightweight, wireless and robust. Since EMG signals alone are complex to classify, the ideal device should include extra sensing modalities. An inertial measurement unit (IMU) is particularly well suited for such requirement. In 2014, an accessible and powerful wireless gesture control armband, called the Myo[1], was released. It includes a nine degree-of-freedom IMU as well as eight electromyogram sensors (EMG) disposed symmetrically around the forearm.

Two experiments involving two dancers were conducted and will be detailed in Section III-A and Section IV respectively. For the first experiment, to achieve a robust and authentic feeling of interaction between a robot and a performer, as well as unlocking the full creative potential of the dancers equipped with the Myo, an initial step of *babbling* is done. During *babbling*, information collected from the Myo is directly shown to the dancer by feeding the signal into motorized light projectors. This grants the performer freedom to improvise complex, aesthetically-pleasing choreographies that are still potentially discernible without the need of deep understanding of the device's sensibilities. The frequency of

[1] https://www.thalmic.com/

the armband for this first experiment was limited (10 Hz) to emulate multiples dancers performing simultaneously which would severely limits the amount of data per performer that a single system could handle. The second was a live classification using two Myo sensors at maximum frequency (50Hz for the IMU and 200Hz for the EMG). The first Myo on the forearm sent EMG and IMU data while the second on the calf only provided IMU data (hardware restriction from employing two Myo simultaneously). In each experiment, the artist created a new movement lexicon without any inputs from the researchers (13 movements for the first, 9 for the second), from which improvised seances (or performances) were produced.

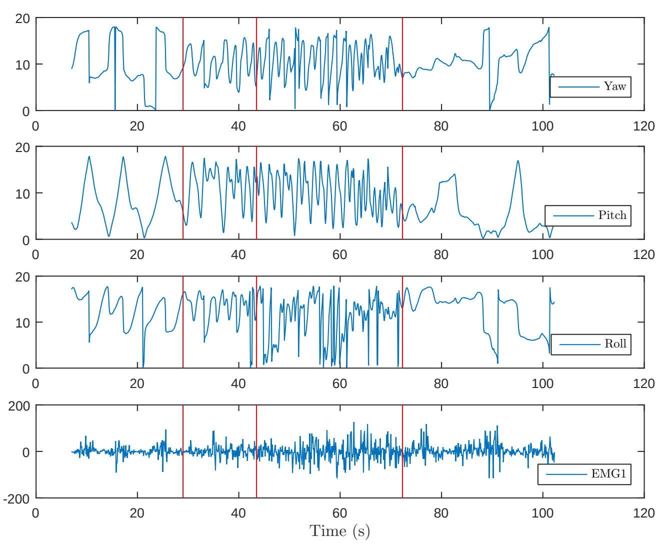

Fig. 2. Example of signals for the EMG and the IMU of the Myo armband, for four different dance movements. Vertical red lines indicate the transition between two movements. The goal of our system is to identify automatically these four movements.

III. OVERALL DESCRIPTION

A key component of this work is the use of machine learning to recognize each movement from the signals captured by the Myo armband. Employing these machine learning techniques allows the system to be robust to uncertainties and noise. Fig. 2 shows a number of such Myo signals displayed at 10Hz, for four different dance movements. Several classifiers were tested in order to select the one achieving the best accuracy while being fast enough to be employed in real-time. The four classifiers were AdaBoost, SVM-RBF, SVM-Linear and Random Forest (RF). They are all well known state-of-the-art learning algorithms.

The rest of this section discusses the created dataset, the principal features used, the selection of hyper-parameters for the various classifiers and how to extract the most meaningful features.

A. Dataset building: Myo low sampling frequency with feedback

For the initial *babbling* step the dancer was assisted by a choreographer, essentially to provide insight on the perception of her movements from an external point of view.

In the end, a first lexicon of 13 moods were created during this experiment. Following this, the dancer performed five improvised performances of three minutes each comprised of a combination of moods from this lexicon which were recorded by the Myo and a camera. The choreographer and dancer then labeled the sequences in each performance using the camera recording. Those labels were then defined as the ground truth. The Myo signals recorded during these three minutes performances in association with the labels constituted this first dataset.

B. Initial Feature Extraction

A state-estimation filter built in the IMU already provides the three orientations of the forearm; its data is thus simple to interpret. As described previously, the raw EMG data is, on the contrary, very noisy and difficult to directly associate with a specific movement. Based on the general similarity between accelerometer data from tactile sensing and the EMG signals of the Myo, some features were borrowed from work pertaining to surface identification [21], to which others from [22], [23] were added. In the end, the following features were extracted from temporal windows of one to three second duration:

- Minimum and maximum values,
- Mean, variance, skewness, kurtosis and the fifth moment.
- Integrated Values (IV - sum of absolute value of the signal),
- Mean Absolute Value (MAV - average of the sum of absolute value of the signal),
- MAV1 (similar to MAV but using weighted average with more weight given on central values),
- MAV2 (as MAV1, with a different weight distribution),
- Root Mean Square (RMS - The square root of the mean square of the signal),
- Zero Crossing (ZC - Number of time the signal cross the value zero),
- Waveform Length (WL - The sum of the difference between two consecutive data point over the signal),
- Slope Sign Change (SSC - The number of change between positive and negative slope of the signal),
- Willison Amplitude (WAMP - number of overreached difference in the signal amplitude), and
- Median Frequency.

Due to limitations of the Myo software, it was not always possible to sample the various signals at a frequency greater than 10Hz in some experimental setup. For this reason, aside from the median frequency, no other frequency-based features were extracted from the EMG signals in the first experiment (Section III-A). As the main spectral contributions of the muscular reactions to the EMG are in the range of 5-10Hz to 400-450Hz [24], it is expected that there will be a reduced classification performance. In the second experimentation, however, frequency-based features are added to the classifier inputs (Section IV) to enhance its performance.

As a clarification, it is important to distinguish between the type of features extracted (e.g. mean, variance, Zero Crossing, etc.) and the features ensemble generated from all these feature types. In fact, there are eight channels from the sEMG and three from the IMU for a total of 11 channels each having 17 features per channel, creating a sub-window of 187 features. The number of sub-windows per example was directly dependent on the length of the window, each sub-window overlapping the previous one by one sample. Hence, for a window of one second, 10 sub-windows were generated whereas for a window of three seconds, 30 were created, representing a total of 1870 and 5610 features per example respectively. Employing multiple sub-windows within a window allowed a measure of the features' variation through time within one example.

C. Classifier and hyper-parameter selection

The classifier was selected based on the results from four different classifiers: RF, support vector machines with two different kernels (Linear and RBF) and AdaBoost. The latter was selected as it is, like RF, an ensemble method. For each algorithm, we repeated the following procedure five times: learning on four artistic performances and testing on the fifth (each time changing the test performance). The reported metric is the average success rate of these five runs. On each run, the hyper-parameters of the algorithms were set by Cross-validation on the four training performances. The hyper-parameter list were respectively:

- RF: the number of estimators and the number of features utilized. The grid search over the number of estimators was 10, 30, 50, 70, 100, 200, 500 and 1000. The maximum number of features could then be equal to $\sqrt{N}$, $log(N)_{base2}$ or the total number of features, where N is the number of estimators.
- SVM-RBF (also designed as SVM-B): the soft margin tolerance hyper-parameter C and the parameters γ (the parameter related to Radial Basis Function RBF kernel) were chosen by grid search with values from 10^{-5} to 10^{5} on a logarithm scale, with 20 values equally scattered per hyperparameters
- SVM-Linear: C was chosen from 10^{-5} to 10^{5} on a logarithm scale with 20 values.
- AdaBoost: the number of estimators was found by grid search from the same group as RF. The learning rate was between 10^{-2} and 1 on a logarithm scale, with 20 values equally distributed on the scale.

All algorithms were taken from `scikit-learn` [25].

The results presented in Fig. 3 show the four classifiers trained with all the features discussed in Sec. III-B with the dataset created in Sec. III-A. In these experiments, the RF was clearly superior. This is expected, as RFs can cope with a large proportion of noisy features in very high dimensional space. Indeed, RFs can handle thousands of features with a relatively small dataset. It also has the capacity to perform well on datasets that are non-linear and where features have a high-level of interaction between them [26]. RF could also be trained, within a minute by a standard laptop. This is a key characteristic in a performance context, as the choreographers need to get feedback rapidly about the reliability of movement detection.

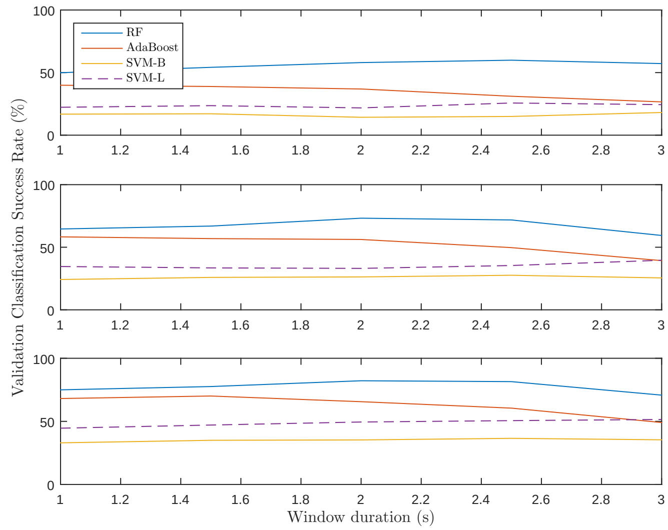

Fig. 3. Validation result comparing RF with AdaBoost, SVM-L and SVM-B. Bottom is the Top3 classification result, then Top2 and Top1. (Higher is better)

D. Automatic Feature Selection

Feature selection was performed automatically, in an iterative manner. We exploited the fact that RF classifiers naturally select the most informative features for classification purposes. Thus at each iteration, the RF classifier with 500 estimators was trained ten times, then the average importance of each feature was computed. During this procedure, the maximum number of features per tree equal to $log(N)$, where N is the number of features of the current iteration. Both values were chosen based on the grid search, described in the previous section. Any feature that was not present amongst the top 20% was rejected, as the RF did not consider them as part of the most relevant group. Features were treated individually for each signal, as the most informative features might vary from one sensor to another. Once every feature not present at the top 20% has been removed, the process iterates. This procedure is repeated until *i)* the accuracy in validation starts to decrease for two consecutive iterations or *ii)* the first 20% feature group has not changed between iterations. At this point, the best feature set is the previous one that provided the highest accuracy in validation. As it was the case in the example of Fig. 4 (one second window), five iterations or less were generally sufficient. In the later example, the final features selected for the RF classifier were:

- for IMU orientation: Maximum, Mean, Variance, IV, MAV and RMS;
- for EMGs: Variance, IV, MAV and RMS.

Other experiments with RF on EMGs suggest to distribute weights on each feature and conserving all of them [27]. Minimizing the features space is more suitable considering

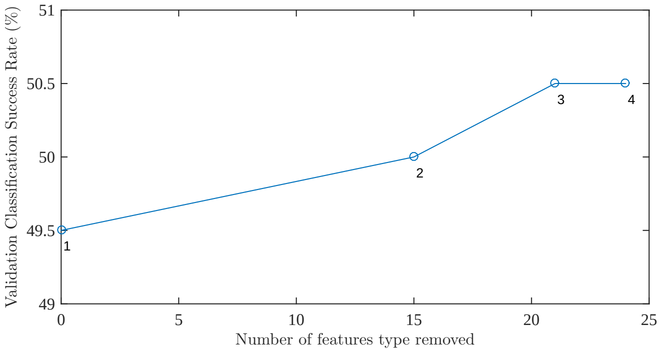

Fig. 4. RF average performance over 10 trainings with features removed and windows of one second, during the automatic feature selection phase. The number next to a data point is the iteration number.

the real-time requirement as doing otherwise would incur too much processing delays.

E. Overlapping Time-window Classification

The tests described in the previous section each employed time-windows ranging from one to three seconds (10 to 30 samples). Notwithstanding transitions between classes, the longer this time-window, the more precise the extracted features are. This however, adds a detrimental latency for the detection of transitions. As the dancers will be performing live interactions with the robots, this latency must be limited. In the comparison presented in Section III-A, the time-windows were non-overlapping, *i.e.* no two samples belonged to the same window.

To increase the training set size (as data is expensive and time-consuming to collect), we tested the use of overlapping windows as a form of data augmentation. New windows were generated by sliding the window by exactly one sample. For a time-window of duration L, this means that $L-1$ samples were shared between consecutive windows. This approach is also minimize the latency of classification during a live performance. These overlapping windows were tested with the RF classifier only, as it was deemed the best in Sec. III-C. Also, only the subset of features determined in Sec. III-D were used. The Top1 and Top3 performances using 2.5 second windows were 72% and 88% respectively. In TopX, if the correct classification is within the X most confident prediction from the classifier then it is characterized as a classification success. Note that a dummy classifier always predicting a random class would achieve a Top1 accuracy of 7.69 %. This first experimentation shows that machine learning techniques allows the recognition of multiples moods with few examples even at low frequency rate. This method would thus be suitable for performances involving multiples dancers with limited computational resources. The next experiments explore the case where higher frequencies and an additional armband are leveraged.

IV. Live experiment with high sampling rate and without feedback

The next set of experiments was conducted to complement several aspects that the previous experimental setup did not

Fig. 5. Dancer Claudia Tremblay during one of the performance.

allow:

- Leveraging the full 200 *Hz* frequency of data coming from the EMG for classification.
- Study of the performances with the same movement classes but different body orientations.
- Employing an additional Myo armband on the calf to extend the possible movements that can be recognized.

The datasets utilized in this section were acquired on a dancer wearing two Myo armbands: one on the forearm and one on the calf. Both IMU orientations were extracted at almost 50 *Hz* and the forearm's eight EMG channels at almost 200 *Hz*. The performer was instructed to develop nine movements that she then regrouped into three different choreographies (Fig. 5) refered to as *moods*. The *moods* were differentiated by the emotion that they each subjectively represented. The training sets were recorded with the performer repeating the nine movements for 20 seconds each. This method of collecting training data offer the advantage of obtaining the ground truth labels without any input from the dancer/choreographer. From this lexicon, the dancer then create three performances of about three minutes each. One of these is serves as a validation set, and the two others to test the performance of the classifier. A fourth performance was created to measure the impact on the classifier's accuracy of variable body orientations with the same lexicon.

A. Classifier performances and feature optimization

The use of two Myos and a higher data acquisition frequency are taken into account and the automatic feature algorithm (Section III-D) is applied on the current dataset. Furthermore, only two of the eight available EMG channels were employed (selected by cross-validation on the validation set). Note that as the IMU and EMG sensors are operating at different sampling rate, the number of data point per sub-window for each modality was set so that they covered the same amount of time. The new subset of selected features for the EMG were: Variance, IV, and MAV. The IMU features resulting from the features selection were: Maximum, Mean, Variance, IV, MAV and RMS. However, as the dancer was executing the movements with slight body orientation variation, RF over-fitted the absolute orientation of the performer using the YAW data. This led to significant drop in accuracy on the validation set. To overcome this, only the variance of the YAW (for both IMU) was employed as it was the only feature that was agnostic to the absolute orientation of the dancer. Furthermore, window of 1 second with an overlap of $0.3s$ were applied to achieve a low-latency classification. The final settings for the classification performances showed in Fig. 7 were:

- Random Forest algorithm with the previously detailed hyper-parameters.
- One second time-window to detect a class.
- Overlaps of $0.3s$ between successive windows.
- Measurements of the orientation of the right forearm and right calf, with two EMG channels on the forearm.

It resulted in a performance of 88% for the nine movements and 94% for the three *moods* (average for both on 100 runs on each performance) with most of the errors arising around the transitions. Additionally, more informal user experiments were conducted with the final system to qualitatively assess the flexibility of the system. In those experiments two professional dancers, one of which is also a professional choreographers analyzed the output of the classifier in real-time (i.e. during the dance performance) and agreed both on the coherence of the predictions and the low latency of the classifier. These performances are thus suitable to close the loop with a robot having a sense of which emotional ambiance the user is in.

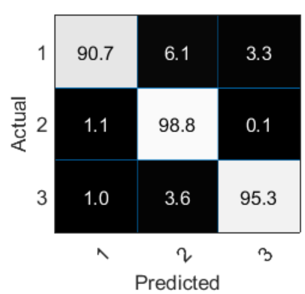

Fig. 6. Confusion matrix (in %) for RF trained on three *moods*.

Training the classifier on the movements that are then mapped to their respective *moods*, yielded a success rate of 95 % for the three *moods* (average on 100 runs on each performance). Fig. 6 shows between which *moods* the few wrongful blends occurred. If it increased the success rate by only 1%, it is relevant to keep this approach in mind for larger lexicons.

B. Impact of IMU vs. EMG data for single and dual armbands

We compared the informative value of the EMG from the arm EMG_a against the IMU of the leg (IMU_l) and arm (IMU_a) using the accuracy of RF on the nine classes. The classification results for all possible combinations is shown in Fig. 7. One can see that the IMU, as a single source of information, performs better than the EMGs, which is not surprising given the complexity and noise of the latter. Taken separately, the IMUs on the forearm and calf give similar performances but together they increase the performance of the classifier by more than 10%. Valuable information can still be gained from the EMG, however. Alone, the forearm Myo is better without EMGs signals; but when combined to the calf and forearm IMU, the two EMGs channels increased the accuracy by up to 3%.

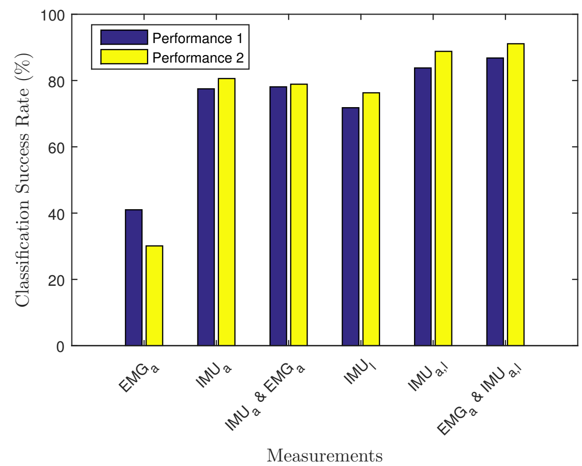

Fig. 7. RF 9-class performances on the second experiment. (a) means arm and (l) means leg.

C. Absolute orientation over-fit

If the performance is not created to face a specific orientation as is the case when performing in theatre-in-the-round (i.e. the audience surrounds the dancer). Care must be taken to include different body orientations when building the training dataset, otherwise the absolute orientation of the dancer during training will be over-fitted by the classifier. A fourth performance in this second experiment was dedicated to test this issue. The dancer was instructed to continuously change her orientation while building a new training dataset of 20 seconds per movements referred to as *orientation training dataset*. The dancer also created a new performance of three minutes, again while continuously changing her body orientation. Apart from this, all the other parameters in this test (e.g. number of training/test examples, hyper-parameters, features) are the same as described earlier in this section. The accuracy on this new performance using the precedent training dataset is 52% and 75% for the nine movements and the three *moods* respectively. However, when using the *orientation training dataset*, the accuracy grows to 68% and 84% respectively which are in a workable range for the use-case described in the article. Because this type of performance introduces a lot more variance between examples of the same gesture, more training data should be collected to achieve classification accuracy similar to those obtained in Sec. IV-A.

V. DISCUSSION AND FUTURE WORKS

Throughout this paper, we explored the scenario of leveraging a wearable sensor device to control a robot in the context of a live dance performance. We demonstrated an end-to-end system that allowed a dancer to create a lexicon of emotionally-driven movements that a robot is able to react to in real-time through our machine learning oriented system. The proposed system employs Myo armbands which have the advantages of being inexpensive, fast to use, non-intrusive and work well even with obstructed line of sight. The signal they produce is however noisy compared to Mocap or camera systems. We showed that our machine learning driven approach nevertheless achieved a robust classifications accuracy in real-time at a frequency of 1 *Hz* of 88% and 94% for nine movements and three *moods* respectively using only 20 seconds of training data per movement. We noted that increasing the size of the windows improved the classifier's success rate, at the expense of its applicability in live settings (due to increased latencies). We also presented that including EMG signals in combination with IMU signals produce a small but noticeable increase in accuracy. Thus, when possible, integrating EMG signals into the classifier should be preferred. However, for performances including multiples dancers, only employing IMU signals can be a good trade-off between the classifier performance and computational load. Amongst the limitations of such a system, we showed sensibility to the body orientation, but still within an acceptable performance. We hypothesized that the body orientation problem could be minimized by collecting more data for the training set in order to augment the accuracy.

Future experiments include the use of the system in a live setting with a robotic device in order to close the loop. Ultimately, the robot determines its movements from the performer's, allowing the emergence of an improvised and hybridized "pas de deux". Adding other interaction interfaces such as singing recognition, the artists will be able to create a hybrid choreography with the machine. We will also explore the problem of body orientation and test if a wider training set would help achieve higher accuracy. Finally, experiments with multiple dancers will also be performed as the proposed system is particularly well suited for this application.

ACKNOWLEDGMENT

The authors would like to acknowledge the financial support of the Fonds Quebécois de Recherche sur la Nature et les Technologies (FQRNT). Special thanks to the performers Claudia Tremblay, Ghyslaine Doté and Tanya Dolbec and to their choreographers Elisabeth Toussain and Armando Menicacci.

REFERENCES

[1] Stelarc, "Third hand," http://stelarc.org/?catID=20265, accessed: 2016-11-29.

[2] S. Wilson, *Information Arts: Intersections of art, science and technology*. Cambridge, Massachusetts: The MIT Press, 2002.

[3] D. St-Onge, N. Reeves, and N. Petkova, "Robot-human interaction: A human speaker experiment," in *Proceedings of 12th Annual Conference on Human-Robot Interaction*, Vienna, Austria, March 2017, pp. 1–8.

[4] D. St-Onge, P.-Y. Brèches, I. Sharf, N. Reeves, I. Rekleitis, P. Abouzakhm, Y. Girdhar, A. Harmat, G. Dudek, and P. Giguère, "Control, localization and human interaction with an autonomous lighter-than-air performer," *Robotics and Autonomous Systems*, vol. 88, pp. –, 2016.

[5] T. R. Song, J. H. Park, S. M. Jung, and J. W. Jeon, "The development of interface device for human robot interaction," in *Proceedings of the International Conference on Control, Automation and Systems*, 2007, pp. 640–643.

[6] J. Aggarwal and Q. Cai, "Human motion analysis: a review," *Computer Vision and Image Understanding*, vol. 73, no. 3, pp. 428–440, 1999.

[7] A. Sanna, F. Lamberti, G. Paravati, and F. Manuri, "A kinect-based natural interface for quadrotor control," *Entertainment Computing*, vol. 4, pp. 179–186, 2013.

[8] K. Petersen, J. Solis, and A. Takanishi, "Musical-based interaction system for the Waseda Flutist Robot: Implementation of the visual tracking interaction module," *Autonomous Robots*, vol. 28, no. 4, pp. 471–488, 2010.

[9] S. Cosentino, K. Petersen, Z. H. Lin, L. Bartolomeo, S. Sessa, M. Zecca, and A. Takanishi, "Natural human-robot musical interaction: understanding the music conductor gestures by using the WB-4 inertial measurement system," *Advanced Robotics*, vol. 28, no. 11, pp. 781–792, 2014.

[10] F. Weichert, D. Bachmann, B. Rudak, and D. Fisseler, "Analysis of the accuracy and robustness of the leap motion controller," *Sensors*, vol. 13, no. 5, pp. 6380–6393, 2013.

[11] R. Herrera-Acua, V. Argyriou, and S. A. Velastin, "A kinect-based 3d hand-gesture interface for 3d databases," *Journal on Multimodal User Interfaces*, vol. 9, no. 2, pp. 121–139, June 2015.

[12] J.-H. Ahn, C. Choi, S. Kwak, K. Kim, and H. Byun, "Human tracking and silhouette extraction for humanrobot interaction systems," *Pattern Analysis and Application*, vol. 12, pp. 167–177, 2009.

[13] K. DoHyung, L. Jaeyeon, Y. Ho-Sub, K. Jaehong, and S. Joochan, "Vision-based arm gesture recognition for a long-range humanrobot interaction," *Journal of Supercomputer*, vol. 65, pp. 336–352, 2013.

[14] C. Salter, *Entangled: Technology and the transformation of performance*. Cambridge, Massachusetts: MIT Press, 2010.

[15] N. Bu, M. Okamoto, and T. Tsuji, "A hybrid motion classification approach for emg-based humanrobot interfaces using bayesian and neural networks," *IEEE Transactions on Robotics*, vol. 25, no. 3, June 2009.

[16] X. Liu, M. Zhang, A. Richardson, T. Lucas, and J. Van Der Spiegel, "The virtual trackpad: an electromyography-based, wireless, real-time, low-power, embedded hand gesture recognition system using an event-driven artificial neural network," *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2016.

[17] P. K. Artemiadis and K. J. Kyriakopoulos, "A switching regime model for the emg-based control of a robot arm," *IEEE Transactions on Systems, Man and Cybernetics–Part B: Cybernetics*, vol. 41, no. 1, February 2011.

[18] M. V. Liarokapi, P. K. Artemiadis, and K. J. Kyriakopoulos, "Task discrimination from myoelectric activity: A learning scheme for emg-based interfaces," in *Proceedings of the IEEE International Conference on Rehabilitation Robotics*, 2013, pp. 1–6.

[19] A. Xu, G. Dudek, and J. Sattar, "A natural gesture interface for operating robotic systems," in *Proceedings of the 2008 IEEE International Conference on Robotics and Automation (ICRA '08)*, Pasadena, California, USA, May 2008, pp. 3557–3563.

[20] L. Blom and L. Chaplin, *The Intimate Act of Choreography*, ser. Pitt paperback. University of Pittsburgh Press, 1982. [Online]. Available: https://books.google.ca/books?id=OBJHNrNk8E0C

[21] P. Giguere and G. Dudek, "A simple tactile probe for surface identification by mobile robots," *IEEE Transactions on Robotics*, pp. 534–544, 2011.

[22] A. Phinyomark, C. Limsakul, and P. Phukpattaranont, "A novel feature extraction for robust emg pattern recognition," *Journal of Computing*, vol. 1, no. 1, pp. 71–80, 2009.

[23] A. Phinyomark, S. Hirunviriya, C. Limsakul, and P. Phukpattaranont, "Evaluation of emg feature extraction for hand movement recognition based on euclidean distance and standard deviation," in *Proceedings of the IEEE International Conference on Computer Telecommunications and Information Technology*, 2010.

[24] R. Merletti, "Standards for reporting emg data," *Journal of Electromyography and Kinesiology*, vol. 9, no. 1:III-IV, February 1999.

[25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and al, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[26] Y. Qi, "Random forest for bioinformatics," *Ensemble Machine Learning*, pp. 307–323, 2012.

[27] B. Xu, J. Z. Huang, G. Williams, Q. Wang, and Y. Ye, "Classifying very high-dimensional data with random forests built from small subspaces," *International Journal of Data Warehousing and Mining*, vol. 8, no. 2, April–June 2012.