Probabilistic Model of Pianists' Arm Touch Movements

Aristotelis Hadjakos **Erwin Aitenbichler** **Max Mühlhäuser**
TU Darmstadt
Department of Computer Science
telis,erwin,max@tk.informatik.tu-darmstadt.de

Abstract

Measurement of pianists' arm movement provides a signal, which is composed of controlled movements and noise. The noise is composed of uncontrolled movement generated by the interaction of the arm with the piano action and measurement error. We propose a probabilistic model for arm touch movements, which allows to estimate the amount of noise in a joint. This estimation helps to interpret the movement signal, which is of interest for augmented piano and piano pedagogy applications.

Keywords: Piano, arm movement, gesture, classification, augmented instrument, inertial sensing.

1. Introduction

Sensor signals of piano playing movements can be used to provide an additional channel of control over live electronics. One approach is to use the sensor signal as a direct input for an electronic effect. A certain amount of uncontrollability, which may eventually be accepted as musically interesting, is present in such a set-up. Our approach is to automatically analyze the movement signal to distinguish between controlled and uncontrolled movements. Controlled movements can then be used to drive the effect.

Movement signals are useful for piano pedagogy applications. Sonification and visualization can be used to increase awareness and knowledge of playing movements. This can have positive effects on the student's technique and confidence through increased feeling of control. A pedagogy system has to ignore movement measurements that are not controllable by the player to be effective.

A basic question for the analysis of touch movements is to determine if the player has used controlled movement in a given joint to execute a touch. One might think that this question can be easily answered by examining the movement signal. However, the signal is often not pure enough as the following example may illustrate: Assume that a pianist executes a touch with the right arm using movement from the elbow joint exclusively. Before the touch movement begins, the thumb, which will press the key, is located, say, 1cm above the key. The player uses controlled movement in the elbow joint to move the thumb towards the key. When the thumb comes in touch with the key, the right forearm starts to passively rotate clockwise. The forearm rotation signal indicates movement although the user has not used controlled forearm rotation movement. This effect occurs before the hammer strikes the string, i.e., during touch execution.

We propose a probabilistic model of arm movement, which allows to estimate the amount of uncontrolled movement in a joint. A function that estimates mean and standard deviation of the uncontrolled movement in a joint is learned from a large data-set of touch movement samples. The amount of uncontrolled movement in a joint is estimated as a function of the measured movements in the other joints. In the above example, the mean and variance of uncontrolled forearm rotation movement changes when the player uses elbow movement.

The probabilistic model can be used in the following way to determine if the player has used controlled movement in a given joint to execute a touch:

1. The movement in all arm joints is measured.

2. The mean and standard deviation of the uncontrolled movement in a joint is estimated through the learned function, which receives the measured movement in all other joints as input.

3. If the measured movement exceeds the mean significantly, taking the estimated standard deviation into account, the touch was executed with controlled movement in the examined joint.

The remaining paper is structured as follows. In section 2 related work is discussed. The proposed method is discussed in sections 3 to 6. In section 7 we describe how to apply the model for inertial sensors. An evaluation is presented in section 8. Conclusions and future work conclude the paper (section 9).

2. Related work

Recognition methods for movements of instrumentalists have been previously proposed, especially for stringed instruments. Peiper et al. use an electromagnetic motion tracking system to capture violin playing movements [6]. Features are extracted from the data and are used for classification based on decision trees. The system visualizes feature and classification data. Rasamimanana et al. use an inertial measurement sensor to capture violin playing movements [7]. Features are extracted from the data and used for classification based on k-nearest-neighbor. Young uses an inertial sensor to capture violin playing movements [8]. The proposed classification method is a combination of principal component analysis and k-nearest-neighbor.

Contrary to methods that we have previously attempted [3, 5], which classify touch movements based on thresholding of features, the boundaries that are defined with the probabilistic model have a physical meaning. Also our previous examinations were limited to specific arm movements, namely movements in the elbow joint [3], and forearm rotation [5]. This paper examines movements in all arm joints excluding movements originating in the shoulder girdle.

3. Feature computation

The human arm has mainly seven degrees of freedom: (1) wrist abduction-adduction, (2) wrist extension-flexion, (3) elbow extension-flexion, (4) elbow pronation-supination, (5) shoulder abduction-adduction, (6) shoulder extension-flexion, and (7) shoulder rotation.

The piano performance is segmented at note-onsets. At each note-onset seven features $(F_1, F_2, ..., F_7)$ are computed for the seven degrees of freedom of the arm. Each feature should be proportional to the amount of movement in the corresponding joint. Additionally we use the feature (F_v), which is proportional to the velocity of the pressed key. Therefore, at each note onset the following vector d, which will be called sample henceforth, is computed:

$$d = (F_1, F_2, F_3, F_4, F_5, F_6, F_7, F_v) \tag{1}$$

4. Model of arm movement

The feature F_i measures the movement M_i in the i-th joint. However, because of inaccuracies of the measurement, F_i is composed of movement M_i and measurement error E_i:

$$F_i = M_i + E_i \tag{2}$$

The measurement error is typically determined by physical effects to a great extent and only to a lesser extent by actual hardware accuracy limitations. It is, e.g., often not feasible to firmly fixate a sensor to the back of the hand so that the sensor exhibits some degree of independent movement.

An important distinction in our model is the distinction between controlled and uncontrolled movement. Controlled movement is intentional movement of the player. Uncontrolled movement is movement that is not under direct control of the player but is instead the result of mechanical interaction with the piano action and biomechanical constraints of the arm (see section 1).

The movement M_i is composed of controlled movement M_{Ci} and uncontrolled movement M_{Ui}. Therefore, the feature F_i is given as:

$$F_i = M_{Ci} + M_{Ui} + E_i \tag{3}$$

We model $M_{Ui} + E_i$ as normally distributed with some mean μ_i and standard deviation σ_i.

$$M_{Ui} + E_i \sim \mathcal{N}(\mu_i, \sigma_i) \tag{4}$$

The mean μ_i and the standard deviation σ_i of the uncontrolled movement and measurement error of the i-th joint $M_{Ui} + E_i$ is a function of the controlled movement in the other joints M_{Cj} with $j \neq i$ and the key velocity F_v.

$$(\mu_i, \sigma_i) = f_i(M_{C1}, ..., M_{Ci-1}, M_{Ci+1}, ..., M_{C7}, F_v) \tag{5}$$

It is necessary to include the key velocity F_v, which provides information about finger movement intensity, because finger movement intensity has an influence on uncontrolled arm joint movements and is not measured by the features F_1 to F_7.

5. Touch analysis

Assume that the function f_i that expresses the relationship of the controlled movement in the other joints M_{Cj} with $i \neq j$ to the mean μ_i and standard deviation σ_i of the uncontrolled movement in the i-th joint is given.

When a note is played, the feature vector $(F_1, F_2, ..., F_7, F_v)$ is computed. To evaluate f_i it is necessary to estimate the controlled movements M_{Cj} with $j \neq i$. By setting the uncontrolled movements and measurement errors of the other joints $M_{Uj} + E_j$ to zero, we can estimate M_{Cj} with the feature F_j and evaluate the function f_i. [1]

$$\begin{aligned}(\mu_i, \sigma_i) &= f_i(M_{C1}, ..., M_{Ci-1}, M_{Ci+1}, ..., M_{C7}, F_v) && (6)\\ &\approx f_i(F_1, ..., F_{i-1}, F_{i+1}, ..., F_7, F_v) && (7)\\ &=: f_i(d) && (8)\end{aligned}$$

6. Learning f

To learn the function f_i, which estimates mean μ_i and standard deviation σ_i of the uncontrolled movement in the i-th joint, it is necessary to obtain a data-set of movement samples without controlled movement in the i-th joint. This is accomplished by recording movements of touches where the joint i is intentionally not used by the player. E.g., to estimate the function f_2, which estimates mean μ_2 and standard deviation σ_2 of the wrist extension-flexion joint, we have to collect movement samples of touches without involvement of wrist extension or flexion.

Let D be our data-set for learning f_i as described above. D is composed of the samples $d_1, d_2, ..., d_N$, which contain

[1] We introduce $f_i(d)$ for notational convenience, d is the sample $d = (F_1, ..., F_7, F_v)$.

the computed features as defined in section 3. We denote the i-th component of a sample d_n as $d_n(i)$. We learn f_i using maximum likelihood estimation. The likelihood of the data-set D given f_i is:

$$p(D|f_i) = \prod_{n=1}^{N} p(d_n|f_i(d_n)) = \prod_{n=1}^{N} \mathcal{N}(d_n(i)|f_i(d_n)) \quad (9)$$

To allow numerical optimization of the likelihood, f_i has to be given in some parametric form (e.g., f_i could be a linear function of the features so that the parameters would be the linear coefficients). A maximum likelihood estimation of f_i is found by maximizing equation 9 over the parameters of f_i. This maximization is conducted by a numerical optimization algorithm. (To avoid underflow problems $\ln(p(D|f_i))$ is maximized instead.)

7. Applying the model to inertial sensing

We applied the probabilistic model to inertial sensing, using MotionNet, our custom-built inertial measurement system. MotionNet is composed of several inertial sensor units that provide gyroscope and accelerometer signals at a rate of 100Hz. A detailed description of our sensing hardware can be found in [4]. The sensor units are worn on the player's upper arm, forearm, and back of the hand.

7.1. Feature computation

The feature computation is based on the gyroscope signals. To calculate the angular rate in a joint, the gyroscope signals of the two connected limbs are used. The features F_1 to F_7 are then computed by taking the mean value of the angular rate in the corresponding joint over 0.08 seconds.

To compute feature F_3 (elbow extension-flexion), the orientation of the sensor on the forearm and the wrist relative to gravity has to be determined. This is done by calculating pitch and roll angles using Kalman filtering to combine gyroscope and accelerometer signals. To compute the features of the shoulder joint (F_5, F_6, and F_7), the gyroscope signals of the upper arm are directly used since there is no sensors attached to the upper body.

7.2. Key velocity estimation

We used a Kawai K–15 ATX acoustical upright piano with MIDI interface. To estimate the mapping from MIDI velocity to the velocity of the key, we attached a MotionNet sensor unit to the back of the hand of a player. Then the player executed touches from the wrist with different loudness. The maximal angular rate was determined for each touch. This angular rate is proportional to the velocity of the key. An exponential function of the form $g(x) = ax^b + c$ was fitted to the recorded data using the least mean squares algorithm (see figure 1). The input x is the MIDI velocity provided by the piano and the output y is the maximal angular rate. By evaluating the learned mapping g for a given MIDI velocity, the feature F_v can be computed as the value estimated by the function $g(x)$, which is proportional to the velocity of the key.

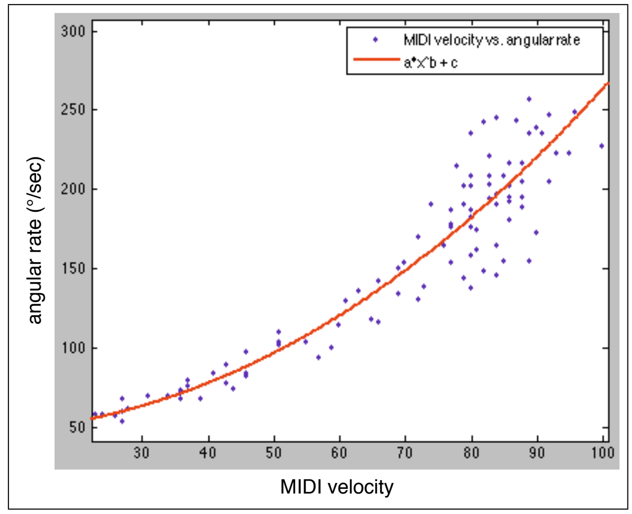

Figure 1. MIDI velocity vs. angular rate.

7.3. Collecting data

The elementary possibilities to execute a touch with the support of the arm are: to move the hand from the wrist (wrist-touch), to move the forearm in the elbow joint (elbow-touch), to rotate the forearm towards the thumb (pronation-touch), to rotate the forearm toward the little finger (supination-touch), and to move the whole arm in the shoulder joint (shoulder-touch).

Our data-set is composed of approximately 18.000 touches, which were executed by one pianist. The recordings were conducted in several sessions. Each session contains approximately 100 touches. Several parameters were identified which have an effect on uncontrolled movements in the pianist's arm joints.

The type of touch has an effect on the uncontrolled movements; e.g., an elbow-touch introduces more uncontrolled movement in the upper arm than a finger-touch. Therefore we recorded different types of touches: finger-touches, wrist-touches, elbow-touches, pronation-touches, supination-touches, and shoulder-touches.

The loudness of a played note has an effect on the uncontrolled movements because of mechanical interaction of the finger with piano action (especially the keybed). Therefore, we recorded the touches in three loudness bands: pianissimo to mezzo-piano, mezzo-piano to mezzo-forte, and mezzo-forte to fortissimo.

The initial height above the key plays a role for the uncontrolled movements, as bigger movements induce more uncontrolled movements in the other joints. Therefore, we recorded direct touches where the execution starts either with the finger on the key and indirect touches where the execution starts with the finger about 1–2 cm above the key.

Which finger executes the touch plays a role for the uncontrolled movements. E.g., if an elbow-touch is executed with the thumb, a supination of the forearm occurs because of the interaction of the arm with the piano action. If an elbow-touch is executed with the middle finger, uncontrolled supination is not typical. Therefore, we recorded touches for each finger.

The recorded touches are grouped into the data-sets D_0 to D_7. The data-set D_i contains all touches with controlled movement in joint i. For example D_2 contains touches with controlled wrist extension-flexion movement, i.e., D_2 contains all wrist-touches. Similarly D_3 contains all elbow-touches, D_4 contains all pronation- and supination-touches, and D_6 contains all shoulder-touches. Additionally D_0 is the data-set that consists of all finger-touches. For the joints $i = 1, 5, 7$ no recordings with controlled movement in these particular joints were made so that $D_1 = D_5 = D_7 = \emptyset$.

The collected data samples originate from a single pianist. Because of differences in anatomy and movement habits, it is optimal to train the estimation functions with data from the specific user of the system only. While such a tailored system is applicable in certain scenarios, e.g., if a pianist records movement samples for her individual augmented piano system, it is also desirable to have an universal estimation function. Therefore, it would be interesting to examine the uncontrolled movements of different pianists in the future.

7.4. Learning f

Many possibilities exist to define the function f_i. We chose to discuss the two functions defined in table 1. Variant 1, the minimalist model, uses solely key velocity. Variant 2, the full model, uses measured movement and key velocity and is motivated below. Variant 1 serves mainly for comparison with variant 2.

To estimate the parameters α_i, β_i, etc., equation 9 is maximized. The data-set that is used to compute the parameters of f_i is the union of the data-sets D_j with $j \neq i$, which are composed of all touches with controlled movement in the j-th joint (see section 7.3). To assure that f_i is always valid, the standard deviation has to be always greater than zero. This is assured by formulating these constraints and using a constraint nonlinear optimization algorithm to maximize the following equation 9 with respect to the constraints.

Variant 1 expresses mean and standard deviation as linear function of the feature F_v. The feature F_v measures the amount of overall movement. The mean is modeled to be linear in the amount of overall movement. The mean is modeled without an intercept term, because it is expected that uncontrolled movement in a joint is very low when overall movement is very low. To assure that the estimated standard deviation is valid, i.e., greater than zero, for all possible velocities F_v, equation 9 is maximized under the constraints $\beta_i \geq 0$ and $\gamma_i > 0$ with a constraint nonlinear optimization algorithm.

Table 1. Variants for mean and standard deviation estimation.

	Variant
1	$\mu_i = \alpha_i F_v$ $\sigma_i = \beta_i F_v + \gamma_i$
2	$\mu_i = \alpha_i F_2^- + \beta_i F_3^- + \gamma_i F_4^p + \delta_i F_4^s + \epsilon_i F_6^- + \zeta_i F_v$ $\sigma_i = \eta_i F_2^- + \theta_i F_3^- + \iota_i \lvert F_4 \rvert + \kappa_i F_v + \lambda_i$

Variant 2 expresses mean and standard deviation as function of measured movement and key velocity. The movements that produce the greatest amount of uncontrolled movements, are movements that can effect a movement of the key. This is because of the interaction with the piano action, which is not present in non-touch-movements like the abduction and adduction of the wrist. Therefore, function 2 expresses the mean and the standard deviation as function of the joint movements that are used for touch movements: wrist flexion F_2^-, elbow extension F_3^-, forearm pronation F_4^p, forearm supination F_4^s, and shoulder extension F_6^-. (The superscript "-" denotes downward movement towards the key.) By definition F_2^-, F_3^-, F_4^p, F_4^s, and F_6^- are always greater or equal to zero. The mean is computed as a linear combination of these features and key velocity, which provides additional information about finger movement. An intercept is not included in the calculation of the mean because the mean of the uncontrolled movement can be expected to be close to zero when there is little overall movement. For the computation of the mean, the forearm rotation feature F_4 is split into F_4^p and F_4^s because the effects of the forearm rotation cannot be modeled as linear in F_4. E.g., when the player executes pronation- or supination-touches, the forearm is lifted by the force of the rotation when the finger hits the keybed. However, this cannot be expressed by a function linear in the forearm rotation movement F_4.

The standard deviation is expressed as a linear function of wrist flexion F_2^-, elbow extension F_3^-, the absolute value of forearm rotation F_4, and shoulder extension F_6-. To compute the standard deviation it is not necessary to split the forearm rotation into pronation and supination because the standard deviation should be similar for similar amounts of pronation or supination.

A constraint nonlinear optimization algorithm is used to find the parameters by maximizing equation 9 under the constraints $\eta_i \geq 0$, $\theta_i \geq 0$, $\iota_i \geq 0$, $\kappa_i \geq 0$, and $\lambda_i > 0$.

8. Evaluation

In this section we want to discuss (1) how well the estimated means and standard deviations fit the observed data, (2) if the inclusion of movement measurement provides an advantage over estimation based on key velocity only, and (3) how well the model performs when used for classification of touch-types.

8.1. Estimation quality

In figure 2 measurements of uncontrolled forearm rotation movement are plotted. The solid lines in figure 2 show the estimated mean (middle line) and standard deviation (outer lines). One can visually convince oneself that the estimated mean and standard deviation are sensible.

In figure 2 all touch types in our data-set are simultaneously plotted. Variant 1 has no information of the arm movement so that it estimates the amount of uncontrolled movement from the key velocity F_v only. It would be favorable, if the estimation of the mean and standard deviation of the uncontrolled movement would adapt to the different conditions of the touch types. That this is in fact accomplished by variant 2 will be shown by examination of the figures 3 and 4.

In figures 3 and 4 the x-axis corresponds to the feature F_v and the y-axis corresponds to the feature F_6, which measures extension-flexion of the arm in the shoulder joint. Because F_v is computed from a discrete MIDI velocity signal, it is possible to split the data into sets of samples with the same value F_v. Each of these sets corresponds to one vertical slice of the graph. For each vertical slice, the average mean and the average standard deviation is computed by evaluating variant 2 of f_i for each sample. The intuition is that, given that we examine only one type of touch alone, the amount of movement from the joint that executes the touch is proportional to key velocity.

The average mean and the mean $\pm$ the standard variation of a slice are marked with thicker dots. By examining the data points and the plotted estimations one can convince oneself that the estimated mean and standard deviation are sensible. Furthermore, variant 2 is adaptive to the type of touch performed, so that it outperforms variant 1, which can only find a global estimation over all touch types. Consider that variant 2 had no direct information about the performed touch but was able to find good estimates of the mean and standard deviation by evaluating the variant 2 of f_i. Note that the mean of uncontrolled shoulder shoulder extension-flexion movement increases (see figure 4). The controlled movement of the forearm in the elbow, generates an uncontrolled movement that moves the upper arm upwards.

Similar plots can be obtained for other joints and other touch-types.

8.2. Classification

The probabilistic model can be used for classification by applying thresholding on the deviation of a measurement from the mean of estimated uncontrolled movement in a joint. If a measurement F_i is more than some constant times the standard deviation away from the estimated mean of uncontrolled movement, then the touch is classified to be a touch with movement in the i-th joint. A touch can be simultaneously assigned to several touch-types, if the measured movement exceeds the expected uncontrolled movement in sev-

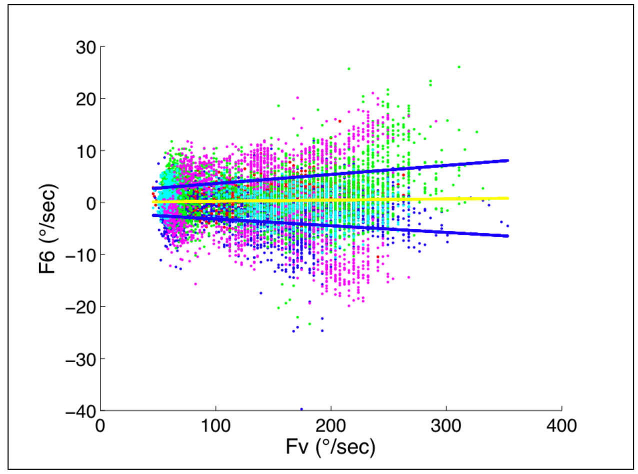

Figure 2. Mean and standard deviation estimation (variant 1).

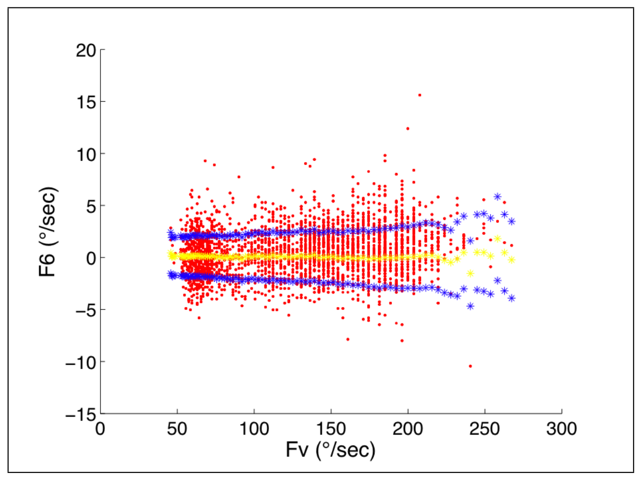

Figure 3. Estimation for finger-touches (variant 2).

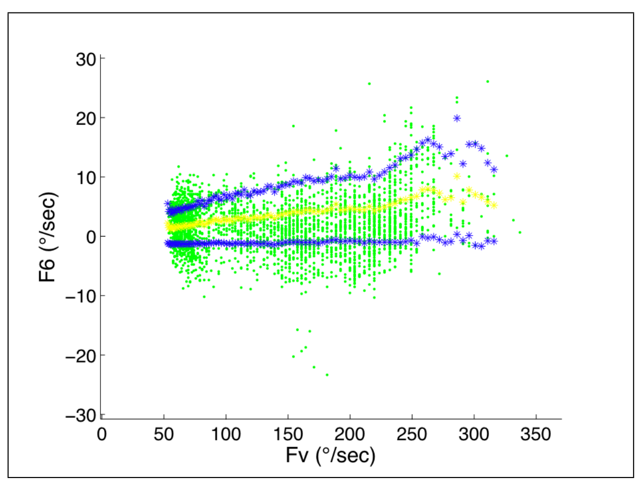

Figure 4. Estimation for elbow-touches (variant 2).

Table 2. Confusion matrix (indications in %).

Class / Actual	Wr.	Elb.	Pro.	Sup.	Sho.
Finger	3.07	3.55	0.11	0.02	0.02
Wrist	**91.78**	3.05	0.16	0.03	0.45
Elbow	5.91	**90.18**	0.09	0.00	0.35
Pronation	0.33	1.33	**99.91**	0.00	0.00
Supination	0.57	4.51	0.00	**99.50**	0.57
Shoulder	0.73	3.92	0.26	0.06	**94.54**

eral joints, which is the case if the pianist executes a touch with controlled movement of more than one joint.

Some types of touch movements exceed the expected amount of uncontrolled movement in the involved joint by a large amount. This is the case for pronation-touches, supination-touches, and to some extent for shoulder-touches. For these separable movements, the classification boundary is given by $\mu \pm 4\sigma$. The controlled movement in wrist- and elbow-touches do not exceed the expected uncontrolled movement by as large amounts. Therefore a classification boundary of $\mu \pm 2\sigma$ is used for classification of wrist- and elbow-touches.

The results of the classification are shown in table 2. Pronation- and supination-touches are recognized with accuracies above 99 percent. Furthermore, other touches are not confused with pronation- or supination-touches.

Other touches are only seldom confused with shoulder-touches. The recognition rate of 94% ensures that a great amount of shoulder touch movements is recognized so that only minute shoulder touch movements are not recognized based on the presented method.

The recognition rate of wrist- and elbow-touches are around 90% so that touches with small amount of elbow or wrist movement are not recognized. Because it is necessary to choose a narrower boundary for classification, in this case $\mu \pm 2\sigma$, other touches are sometimes wrongly classified as wrist- or elbow-touches. An analysis shows that the wrist and elbow movement is in fact not perfectly separable on the given features. E.g., there is a considerable overlap of wrist flexion when comparing finger-touches with wrist-touches with identical key velocities, so that ideal results are not achievable with the given features. Eventually, a different sensor technology or a improved features could provide measurements of wrist and elbow movement with less noise and allow a better classification accuracy.

9. Conclusions and future work

The probabilistic model of arm touch movements allows to estimate the amount of uncontrolled movement and measurement error in a given joint. The mean and standard deviation of the uncontrolled movement in a joint is calculated as a function of the measured movements in the other joints and the key velocity. This function is learned through maximum likelihood estimation from a large data-set of touch movement samples. By applying the probabilistic model to inertial sensing, we were able to show that the model can be successfully used to estimate mean and standard deviations of joint movements. A classification of touch-types based on the estimated amount of touch movements provides very good results for pronation- and supination-touches, good results for shoulder-touches, and reasonable results for wrist- and elbow-touches, which indeed cannot be perfectly separated on the given features.

One advantage of using the probabilistic model for classification is that the classification boundary has a physical meaning: The method determines that a touch has controlled movement in the i-th joint if the measured movement exceeds the estimated uncontrolled movement, i.e., if $|F_i - \mu_i| > a \cdot \sigma_i$. It is therefore possible to weigh the avoidance of false positive classifications with the response to minute movements by changing the constant a. A second advantage of the proposed classification method is that a touch can be classified several touch types, which occurs when the player uses controlled movement in several joints to execute a touch.

The proposed model is not specific to inertial sensing. The use of a different sensor technology, e.g., high-accuracy visual or magnetic tracking could help to increase accuracy. This applies also to a more complete sensor setup where finger movements are also measured.

To build an augmented piano application based on our analysis method, it is necessary to assign played notes to hands. This could be achieved using Computer Vision to track hand movement, e.g., [2].

References

[1] C. M. Bishop, *Pattern Matching and Machine Learning*, Springer, 2006.

[2] D. O. Gorodnichy, and A. Yogeswaran, “Detection and tracking of pianist hands and fingers,” in *CRV’06*, 2006.

[3] A. Hadjakos, E. Aitenbichler, and M. Mühlhäuser, “The Elbow Piano: Sonification of Piano Playing Movements,” in *NIME 2008 Proceedings*, 2008.

[4] A. Hadjakos, E. Aitenbichler, and M. Mühlhäuser, “SYSSOMO: A Pedagogical Tool for Analyzing Movement Variants between Different Pianists,” in *Enactive08*, 2008.

[5] A. Hadjakos, E. Aitenbichler, and M. Mühlhäuser, “Aiding the development of piano gesture recognition with a gesture enriched instrument,” in *Gesture Workshop 2009*, 2009.

[6] C. Peiper, D. Warden, and G. Garnett, “An Interface for Real-Time Classification of Articulation Produced by Violin Bowing,” in *NIME-03 Proceedings*, 2003.

[7] N. H. Rasamimanana, E. Emmanuel Fléty, and F. Bevilacqua. “Gesture Analysis of Violin Bow Strokes,” in *Gesture in Human-Computer Interaction*, Lecture Notes in Computer Science, Springer, 2006.

[8] D. Young, “Classification of Common Violin Bowing Techniques Using Gesture Data from a Playable Measurement System,” in *NIME 2008 Proceedings*, 2009.