Recognizing Hand and Finger Gestures with IMU based Motion and EMG based Muscle Activity Sensing

Marcus Georgi, Christoph Amma and Tanja Schultz

Cognitive Systems Lab, Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Karlsruhe, Germany

Keywords: Wearable Computing, Gesture Recognition, Inertial Measurement Unit, Electromyography.

Abstract: Session- and person-independent recognition of hand and finger gestures is of utmost importance for the practicality of gesture based interfaces. In this paper we evaluate the performance of a wearable gesture recognition system that captures arm, hand, and finger motions by measuring movements of, and muscle activity at the forearm. We fuse the signals of an Inertial Measurement Unit (IMU) worn at the wrist, and the Electromyogram (EMG) of muscles in the forearm to infer hand and finger movements. A set of 12 gestures was defined, motivated by their similarity to actual physical manipulations and to gestures known from the interaction with mobile devices. We recorded performances of our gesture set by five subjects in multiple sessions. The resulting datacorpus will be made publicly available to build a common ground for future evaluations and benchmarks. Hidden Markov Models (HMMs) are used as classifiers to discriminate between the defined gesture classes. We achieve a recognition rate of 97.8% in session-independent, and of 74.3% in person-independent recognition. Additionally, we give a detailed analysis of error characteristics and of the influence of each modality to the results to underline the benefits of using both modalities together.

1 INTRODUCTION

Mobile and wearable computing have become a more and more integral part of our everyday lives. Smartwatches and mixed-reality-glasses are getting popular and widely available, promoting the idea of an immersive usage with micro-interactions. The interaction with such devices differs from the interaction with mobile phones and tablet computers, that already gained functionality that allows to use them as a replacement for conventional computers in a wide range of usage scenarios. For glasses and watches, the usage of onscreen keyboards becomes cumbersome, if not impossible. Therefore, alternative interaction paradigms have to be used, allowing an intuitive handling of these devices.

As gestures performed with the hand and fingers can resemble actual physical manipulations connected to spatial tasks, like navigation on a map or manipulation of a picture, they are a beneficial complementary modality to speech recognition, as these are tasks not easily solved using only spoken language (Oviatt, 1999). A system using hardware that can be worn like clothing or an accessory would be favourable for mobile usage.

Different approaches were proposed to sense both hand and finger movements in a mobile environment without placing sensors directly at the hand of a uesr. They were based on the usage of cameras, body-worn (Mistry et al., 2009) or wrist-worn (Kim et al., 2012), on the measurement of tendon movement (Rekimoto, 2001) or on the usage of Electromyography (EMG) (Saponas et al., 2008; Wolf et al., 2013; Samadani and Kulic, 2014; Kim et al., 2008) and Inertial Meaurement Units (IMUs) (Amma et al., 2014; Hartmann and Link, 2010; Cho et al., 2004; Benbasat and Paradiso, 2001).

This paper presents a recognition framework for gesture interfaces, based on Electromyography and an Inertial Measurement Unit, both being wearable sensor systems.

We will systematically evaluate the performance of this system in differentiating between gestures, using the IMU and EMG individually, as well as the multimodal recognition performance. Additionally, the contributions of both modalities to the overall results will be presented, with focus on the benefits for specific types of movement. This will clarify the advantages and disadvantages of IMUs, EMG, and will validate their combined usage for gesture recognition.

The performance of this system will be evaluated both for session-independent and person-independent

Georgi M., Amma C. and Schultz T..
Recognizing Hand and Finger Gestures with IMU based Motion and EMG based Muscle Activity Sensing.
DOI: 10.5220/0005276900990108
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing* (BIOSIGNALS-2015), pages 99-108
ISBN: 978-989-758-069-7

Table 1: Comparison of investigated research questions. This table shows what kind of gesture recognition tasks were evaluated in the work related to this paper, and which modalities were used.

	session dependent			session independent			person-independent		
	EMG	IMU	combined	EMG	IMU	combined	EMG	IMU	combined
(Li et al., 2010)	-	-	-	-	-	x	-	-	-
(Wolf et al., 2013)	-	-	x	-	-	-	-	-	-
(Zhang et al., 2011)	x	x	x	-	-	-	-	-	x
(Chen et al., 2007)	x	x	x	-	-	-	-	-	-
(Saponas et al., 2008)	-	-	-	x	-	-	x	-	-
(Kim et al., 2008)	x	-	-	-	-	-	x	-	-
(Samadani and Kulic, 2014)	-	-	-	-	-	-	x	-	-
This paper	-	-	-	x	x	x	x	x	x

recognition, which is of importance for practical and usable gesture based interfaces. Session-independent recognition surveys how well the system can be attuned to a user. A system with high session-independent accuracy can be used by one person without training it each time it is used. This makes the system ready to use by just mounting the sensors and starting the recognition. Person-independent recognition is relevant in regard to an envisioned system that can be used without prior training by some new user. Instead, it would use exemplary training data from a selected, representative group, making explicit training sessions superfluous. Such a system could nonetheless still benefit from further adaption to an individual user.

The evaluations of this paper will be based on gestures that were designed to resemble actual physical manipulations, as well as gestures known from the interaction with mobile devices. They were not designed to be optimally distinguishable, but rather to represent useful gestures for real-world interfaces.

The recorded dataset will be made publicly available to be used by others as a common ground for the development and evaluation of gesture recognition systems based on IMUs and EMG.

1.1 Related Work

The simultaneous usage of an IMU and EMG for gesture based interfaces was also evaluated in a small set of other studies, that are listed and compared in Table 1. The comparison of the various different approaches to the task of gesture recognition based only on reported recognition accuracies is hard, as different gesture sets were designed and a variety of recording setups and hardware was used.

In (Li et al., 2010) IMUs and EMG are used for the automatic recognition of sign language subwords, specifically of Chinese sign language (CSL) subwords. On a vocabulary of 121 subwords a high accuracy of 95.78% is achieved, but only session-independent accuracy was evaluated. Additionally, it is hard to estimate the transferability of these results to other tasks, as a very task specific recognizer design is used, in contrast to a more general design, as it is used in this paper and most other studies and most other studies.

In (Wolf et al., 2013) two different approaches for gesture recognition are presented. One uses a SVM to distinguish between 17 static gestures with 96.6% accuracy. It discriminates these 17 gesture classes based on EMG and uses an additional orientation estimation based on IMU signals to distinguish whether the gestures were performed with a hanging, raised or neutral arm, increasing the number of classes that can be interpreted by a factor of three. To decode dynamic gestures, they combine the features of both modalities to a single feature vector. Again, they achieve a high accuracy of 99% for a set of nine gestures. However, both session-independent and person-independent performance are not evaluated for both techniques, and the contribution of the individual modalities is not discussed.

Zhang et al. (Zhang et al., 2011) show, that it is possible to construct a robust, person-independent, gesture based interface using an IMU and EMG to manipulate a Rubik's Cube. They report a person-independent accuracy of 90.2% for 18 different gesture classes. One has to note that this gesture set was designed to include only three different static hand postures that were hard to detect with only an IMU, and that, similar to (Wolf et al., 2013), the direction of movement with the whole arm then increased the number of gestures by a factor of six. In comparison, we use a set of twelve different gestures, with dynamic postures and movement. Additionally, they discuss the contribution of the individual modalities IMU and EMG in a CSL recognition task similar to (Li et al., 2010), but do this only for session-dependent recognition.

Chen et al. (Chen et al., 2007) also compare the session-dependent gesture recognition performance for various gesture sets when using 2D-accelerometers and two EMG channels individually as well as in combination. When using both modalities combined, they report accuracy improvements of $5-10\%$ on various gesture sets with up to 12 classes. Again, this evaluation is not done for the session- and person-independent case.

As we evaluate single-modality performance for person-independent recognition, we also evaluate the performance when using only EMG. Only very few studies have reported on this so far (Saponas et al., 2008; Kim et al., 2008; Samadani and Kulic, 2014), often with lacking results, or only on small gesture sets.

Samadani and Kulić (Samadani and Kulic, 2014) use eight dry EMG sensors in a prototypic commercial armband from Thalmic Labs[1] to capture EMG readings from 25 subjects. In person-independent recognition, they achieve 49% accuracy for a gesture set with 25 gestures. Additionally, they report 79%, 85%, and 91% accuracy for select gesture sets with 10, 6, and 4 gestures, respectively.

2 DATASET

2.1 Gesture Set

To get a reliable performance estimate for practical interfaces, we define a set of gestures that is a reasonable choice for gesture based interfaces. Different studies, e.g. (Alibali, 2005; Hauptmann and McAvinney, 1993), show that spatial gestures can convey unique information not included in linguistic expressions when used to express spatial informations. Therefore we use spatial gestures, as an interface based on such gestures could be beneficial for mobile and wearable usage when used complementary to speech recognition.

Hauptmann et al. (Hauptmann and McAvinney, 1993) evaluate what kind of gestures were performed by subjects trying to solve spatial tasks. In a Wizard of Oz experiment they collected statistical values of the amount of fingers and hands that were used to solve tasks related to the graphical manipulations of objects.

We compared their statistical values with gestures commonly occurring during the actual manipulation of real objects or whilst using touchscreens or touchpads. On the one hand, we hope that users can intuitively use these gestures due to both their level of familiarity, as well as their similitude to physical interactions. On the other hand, we assume that interfaces like virtual or augmented reality glasses might benefit from such gestures by allowing a user to manipualte displayed virtual objects similar to real physical objects.

As a result we defined the list of gestures in Table 2 to be the list of gestures to be recognized. These gestures involve both movements of the fingers, as well as of the whole hand.

2.2 Experimental Procedure

2.2.1 Hardware

Inertial Measurement Unit

A detailed description of the sensor we use to capture acceleration and angular velocity of the forearm during the experiments can be found in (Amma et al., 2014). It consists of a 3D accelerometer, as well as a 3D gyroscope. We transmit the sensor values via Bluetooth and sample at 81.92 Hz.

EMG Sensors

To record the electrical activity of the forearm muscles during the movements of the hand, two *biosignalsplux* devices from Plux[2] are used. These devices allow the simultaneous recording of up to eight channels simultaneously per device. Both devices are synchronized on hardware level via an additional digital port. One additional port is reserved to connect a reference or ground electrode. Each bipolar channel measures an electrical potential using two self-adhesive and disposable surface electrodes. The devices sample at 1000 Hz and recorded values have a resolution of 12 bit.

2.2.2 Subject Preparation

Each subject that was recorded received an explanation of what was about to happen during the experiments. Afterwards 32 electrodes were placed in a regular pattern on both the upper and lower side of their forearm. This pattern can be seen in Figure 1. The position of the electrodes near the elbow was chosen, as the flexors and extensors for all fingers apart from the thumb are mainly located in this area. (*M. extensor digitorum communis, M. extensor digiti minimi, M. extensor pollicis brevis, M. flexor digitorum profundus* and *M. flexor digitorum superficialis*). These muscles are oriented largely parallel to the axis between elbow and wrist. Therefore we applied the electrodes as eight parallel rows around the arm, each row consisting of four electrodes. From each row, the two

[1] Thalmic Labs Inc., www.thalmic.com

[2] PLUX wireless biosignals S.A., www.plux.info

Table 2: All gestures defined to be recognized.

#	Gesture Name	Description	Interaction Equivalence
1	Flick Left (FL)	Flick to the left with index finger extended.	Flicking left on a touchscreen.
2	Flick Right (FR)	Flick to the right with index finger extended.	Flicking right on a touchscreen.
3	Flick Up (FU)	Flicking upwards with index and middle finger extended.	Scrolling upwards on a touchscreen.
4	Flick Down (FD)	Flicking downwards with index and middle finger extended.	Scrolling downwards on a touchscreen.
5	Rotate Left (RL)	Grabbing motion followed by turning the hand counterclockwise.	Turning a knob counterclockwise.
6	Rotate Right (RR)	Grabbing motion followed by turning the hand clockwise.	Turning a knob clockwise.
7	Flat Hand Push (PSH)	Straightening of the hand followed by a translation away from the body.	Pushing something away or compressing something.
8	Flat Hand Pull (PLL)	Straightening of the hand followed by a translation towards the body.	Following the movement of something towards the body.
9	Palm Pull (PLM)	Turning the hand whilst cupping the fingers followed by a translation towards the body.	Pulling something towards oneself.
10	Single Click (SC)	Making a single tapping motion with the index finger.	Single click on a touchscreen.
11	Double Click (DC)	Making two consecutive tapping motions with the index finger.	Double click on a touchscreen.
12	Fist (F)	Making a fist.	Making a fist or grabbing something.

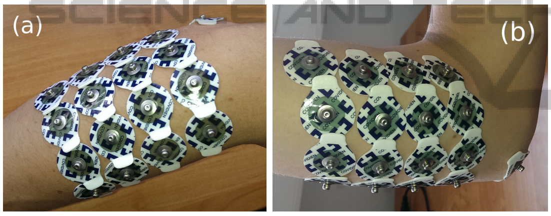


Figure 1: Electrode placements on the (a) upper and (b) lower side of the forearm.

upper, as well as the two lower electrodes form a single, bipolar channel. With eight rows and two channels per row we get 16 EMG channels in total. The reference electrodes for each device were placed on the elbow.

We decided to place the electrodes in a regular pattern around the forearm instead of placing them directly on the muscles to be monitored. This resembles the sensor placement when using a wearable armband or sleeve like in (Wolf et al., 2013) or (Saponas et al., 2008), that can easily be worn by a user. This kind of sensor placement does lead to redundant signals in some channels, as some muscles are recorded by multiple electrodes. However, it should prove useful for future work on how to further improve the session- and person-independent usage, as it allows for the compensation of placement variations using virtual replacement strategies or some blind source separation strategy. Placement variations can hardly be avoided with sensor sleeves that are to be applied by a user.

The inertial sensor device was fixed to the forearm of the subject using an elastic wristband with Velcro.

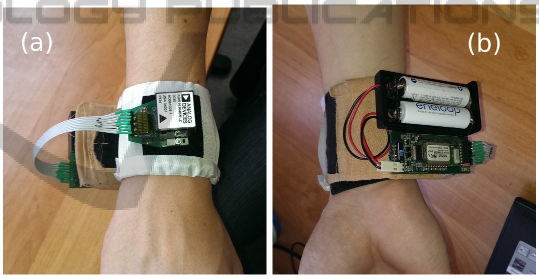


Figure 2: Inertial sensor device mounted ath the wrist on top of the forearm (a). Power supply and controller board are underneath the arm (b).

The sensor itself was mounted on top of the forearm, which can be seen in Figure 2. No hardware is placed on the hand and fingers themselves, all sensing equipment is located at the forearm.

The whole recording setup is not intended to be practical and wearable, but can be miniaturized in the future. IMUs can already be embedded in unobtrusive wristbands and (Wolf et al., 2013) show, that the manufacturing of armbands with integrated EMG sensor channels is possible. Also the upcoming release of the *Myo* from Thalmic Labs shows the recent advances in the development of wearable sensing equipment.

2.2.3 Stimulus Presentation and Segmentation

To initiate the acquisition of each gesture, a stimulus is presented to a subject. Figurative representations for all the gestures defined in Table 2 were created. Additionally a short text was added, describing

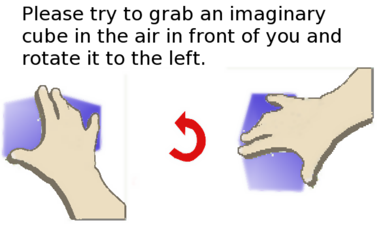

Please try to grab an imaginary cube in the air in front of you and rotate it to the left.

Figure 3: Example of a gesture stimulus, particularly the stimulus of the *rotate left* gesture.

the movements of the gesture. An example of such a stimulus can be seen in Figure 3.

The start and end of each gesture acquisition have to be manually triggered by the subjects by keypress to generate the segmentation groundtruth.

2.3 Data Corpus

In one recording session, 15 repetitions of the twelve defined gestures were recorded in random, but balanced order. We do not use a fixed order of gestures to force the subjects to comprehend and perform each gesture individually, rather than to perform a repetitious pattern. Additionally, this makes the movements of a gesture less dependent on the context of the gestures prior to and after it.

Five different subjects were asked to participate in such recording sessions. Their age was between 23 and 34; four of them were male, one was female. For each subject, five sessions were recorded on different days. Therefore, each gesture was recorded 75 times by each subject, which sums up to 900 gesture performances per subject. In total, 25 sessions with 4500 gesture repetitions are present in the data corpus. The corpus also contains the data samples we recorded during the relaxation phases, as well as the data samples that were transmitted between the actual gesture acquisitions. It might therefore also be used in future work to investigate other tasks and topics, like automatic segmentation, gesture sequence recognition, or the effects of muscle fatigue and familiarization.

3 GESTURE RECOGNITION

3.1 Preprocessing and Feature Extraction

For preprocessing, we normalize the signals of both sensors using Z-Normalization. For the IMU, this normalization decreases the impact of movement speed on the recordings. Furthermore, it reduces the influence of gravitation on the accelerometer signals, that we assume to be a largely constant offset, as the normalization removes baselineshifts. It is characteristic for EMG signals to fluctuate around a zero baseline. Therefore the mean of the signal should already be almost zero and mean normalization only removes a shift of the baseline. But the variance normalization has a rather large impact, as it makes the signal amplitude invariant to the signal dampening properties of the tissue and of skin conductance, up to a certain degree.

After preprocessing, we extract features on sliding windows for both modalities. As the different sensor systems do not have the same sampling rate, we choose the number of samples per window in accordance to the respective sampling rate, so that windows for both modalities represent the same period of time. This allows for the early fusion of the feature vectors for each window to one large feature vector.

Similar to (Amma et al., 2014), we use the average value in each window of each IMU channel as a feature. As the average computation for each window is effectively a smoothing operation and could therefore lead to information loss, we added standard deviation in each window as a second feature for each channel of the IMU.

We also compute standard deviation as a feature on each window of each EMG channel. (Wolf et al., 2013) state that, like Root Mean Square (RMS), standard deviation is correlated to signal energy, but not as influenced by additive offsets and baselineshifts.

Averaging operations on large windows often reduce the influence of outliers by smoothing the signal, whereas feature extraction on small windows increases the temporal resolution of the feature vectors. As it is hard to predict the influence of different window sizes on the recognition results, we evaluated different window sizes in the range of 50 ms to 400 ms. We did not evaluate longer windows, as some of the recorded gestures were performed in under 600 ms. The mean duration of the gestures is about 1.1 s. We got the best results for windows with a length of 200 ms and chose an overlap of half a window size, namely 100 ms.

In conclusion we compute for each window a 28 dimensional feature vector. The first twelve dimensions are mean and standard deviation for each of the six channels of the IMU. The remaining 16 dimensions are the standard deviation of each EMG channel. For single modality evaluations, we only use the features of one modality whilst omitting the features of the other one.

Figure 4 shows the set of features computed for

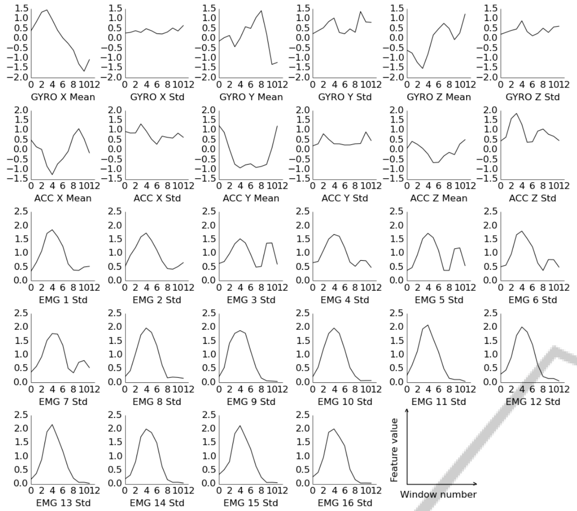


Figure 4: Computed features for the gesture *fist*. All subplots represent one of the 28 feature vector dimensions. The raw signals were z-normalized and windowed into windows of 200 ms length with an overlap of 100 ms between consecutive windows. On each window one value for each dimension was computed. The first twelve subplots show features for the IMU channels, namely mean and standard deviation for the accelerometers and gyroscopes. The remaining 16 subplots show standard deviation for each of the EMG channels.

the signals recorded when making a fist. Each subplot shows the progression of values of one dimension of the feature vector over all extracted windows.

3.2 Gesture Modeling

In our system, continuous density Hidden Markov Models (HMMs) are used to model each of the gesture classes. For an introduction to HMMs refer to (Rabiner, 1989). We chose linear left-to-right topologies to represent the sequential nature of the recorded signals and Gaussian Mixture Models (GMMs) to model the observation probability distribution. Empirically we found topologies with five states and GMMs with five Gaussian components to deliver the best performance. The Gaussian components have diagonal covariance matrices to avoid overfitting to the data, as the number of free model parameters is then only linearly dependent on the number of features, in contrast to the quadratic dependence for full matrices. Provided a sufficiently large dataset, in future work the usage of unrestricted covariance matrices might further improve the gesture modeling by representing the correlation between the different channels. To fit the models to the training data, we initialize them using kMeans and use Viterbi Training afterwards.

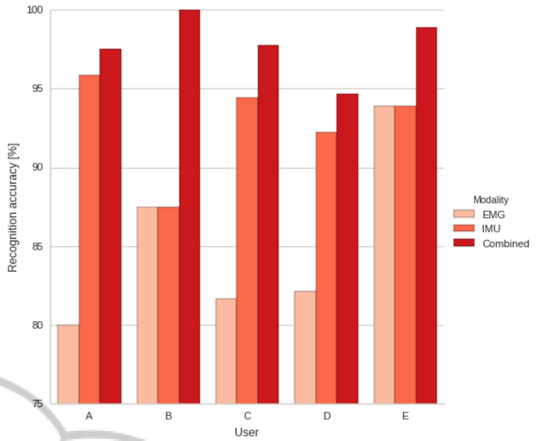


Figure 5: Session-independent recognition accuracy for all recorded subjects. For each subject, the performance when using IMU or EMG individually, as well as in combination, are shown.

4 RESULTS

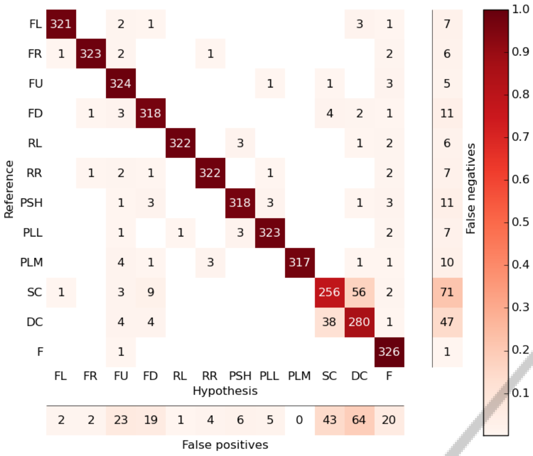
We evaluate the performance of our gesture recognition system for session-independent and person-independent recognition by determining its accuracy in discriminating the twelve different gesture classes of Table 2, with chance being 8.33%. The gesture labels used in the graphics of this section follow the short names introduced in Table 2.

4.1 Session Independent

For the session-independent evaluation, testing is done using cross-validation individually for each subject. The training set for each validation fold consists of all sessions but one of a subject. The remaining session is used as the test set. We achieve 97.8%(±1.79%) recognition accuracy as an average for all evaluated subjects.

Figure 5 displays the individual recognition performance for each subject. With all of the five subjects achieving more than 94% accuracy, the session independent recognition yields very satisfying results. Additionally, Figure 5 shows the recognition accuracy when modalities are used individually. This will be discussed later on.

Figure 6 shows the confusion matrix for the different gesture classes. It has a strong diagonal character, which is fitting the high overall accuracy. A block formation at the crossing between the rows and columns corresponding to *single click* and *double click* illustrates that these gestures are occasionally confused with each other. Only one small movement of the



	FL	FR	FU	FD	RL	RR	PSH	PLL	PLM	SC	DC	F	False negatives
FL	321		2	1							3	1	7
FR	1	323	2			1						2	6
FU			324					1		1		3	5
FD		1	3	318						4	2	1	11
RL					322		3				1	2	6
RR		1	2	1		322		1				2	7
PSH			1	3			318	3			1	3	11
PLL			1		1		3	323				2	7
PLM			4	1		3			317		1	1	10
SC	1		3	9						256	56	2	71
DC			4	4						38	280	1	47
F			1									326	1
False positives	2	2	23	19	1	4	6	5	0	43	64	20	

Figure 6: Confusion matrix for session-independent gesture recognition. The matrix shows what hypotheses were given for different gestures that were fed to the recognizer. The labels follow Table 2.

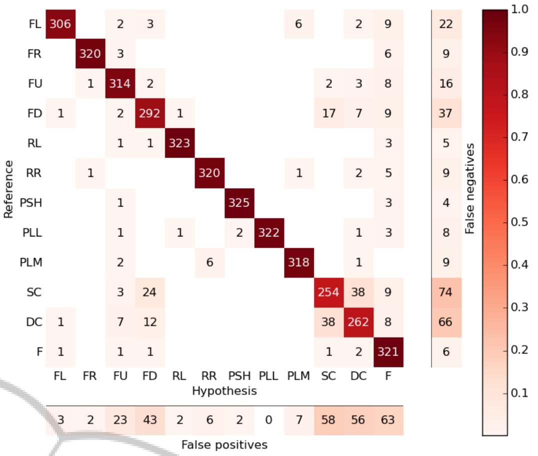
index finger differentiates both gestures. This could easily be confused with involuntary finger movements at the beginning and the end of the gesture recording.

4.1.1 Modality Comparison

Figure 5 visualizes the recognition performance when we use only the feature subset of one of the sensors for the recognition. One can see that the accuracy is lower when the sensing modalities are used individually. Nonetheless rather high results were achieved using both modalities separately, on average 92.8%(±2.88%) with the IMU and 85.1%(±5.09%) with the EMG sensors. Therefore both of the systems are suitable choices for hand gesture recognition.

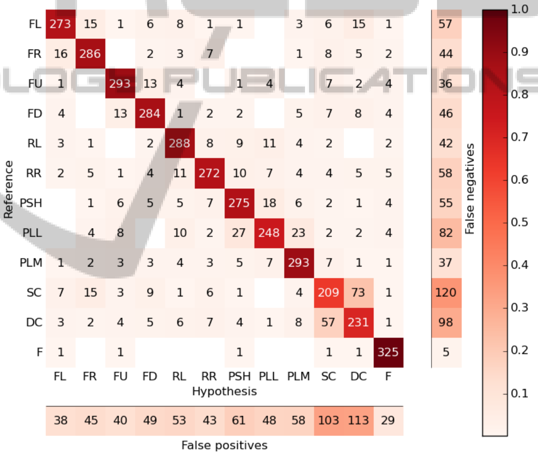
To further analyze the advantages and disadvantages of the IMU and EMG, Figure 7 and Figure 8 show the confusion matrices for session-independent recognition using the individual modalities. We expect the IMU and EMG to be complementary and to differ in their performance for certain gesture classes, due to them monitoring different aspects of the performed movements.

Figure 7 shows the confusion matrix for single modality recognition using the IMU. It is overall rather similar to the one for multi modality recognition in Figure 6. The most prominent difference is, that more false positives were reported for *single* and *double click*, and for *fist*. Whilst the *clicks* are often confused with one another, *fist* is more often given as a hypothesis for the *flick* and *rotate* classes. As making a fist only involves minor arm movements, we assume that the gesture has largely unpredictable IMU features. Thus the HMM for *fist* is less descriptive.



	FL	FR	FU	FD	RL	RR	PSH	PLL	PLM	SC	DC	F	False negatives
FL	306		2	3					6		2	9	22
FR		320	3									6	9
FU		1	314	2						2	3	8	16
FD	1		2	292	1					17	7	9	37
RL			1	1	323							3	5
RR		1				320			1		2	5	9
PSH			1				325					3	4
PLL			1		1		2	322			1	3	8
PLM			2			6			318		1		9
SC			3	24						254	38	9	74
DC	1		7	12						38	262	8	66
F	1		1	1						1	2	321	6
False positives	3	2	23	43	2	6	2	0	7	58	56	63	

Figure 7: Confusion matrices for session-independent gesture recognition. The recognizer used only the features extracted from the IMU signals. The labels follow Table 2.



	FL	FR	FU	FD	RL	RR	PSH	PLL	PLM	SC	DC	F	False negatives
FL	273	15	1	6	8	1	1		3	6	15	1	57
FR	16	286		2	3	7			1	8	5	2	44
FU	1		293	13	4		1	4		7	2	4	36
FD	4		13	284	1	2	2		5	7	8	4	46
RL	3	1		2	288	8	9	11	4	2		2	42
RR	2	5	1	4	11	272	10	7	4	4	5	5	58
PSH		1	6	5	5	7	275	18	6	2	1	4	55
PLL		4	8		10	2	27	248	23	2	2	4	82
PLM	1	2	3	3	4	3	5	7	293	7	1	1	37
SC	7	15	3	9	1	6	1		4	209	73	1	120
DC	3	2	4	5	6	7	4	1	8	57	231	1	98
F	1		1				1			1	1	325	5
False positives	38	45	40	49	53	43	61	48	58	103	113	29	

Figure 8: Confusion matrices for session-independent gesture recognition. The recognizer used only the features extracted from the EMG signals. The labels follow Table 2.

This is consistent with the expectation, that gesture recognition based solely on IMU signals might prove problematic for small-scale hand gestures that do not involve arm movements.

In contrast, one would expect movements involving large rotational or translational movements of the whole arm to be easily discriminated. Accordingly, *rotate right* and *rotate left*, as well as *flat hand push* and *flat hand pull* have a low false positive count.

Figure 8 shows the confusion matrix when using only EMG features. As for the multimodal case, we see most of the confusions for the *double* and *single click* gestures.

Also some other less frequent misclassifications occur when using EMG individually. As we antici-

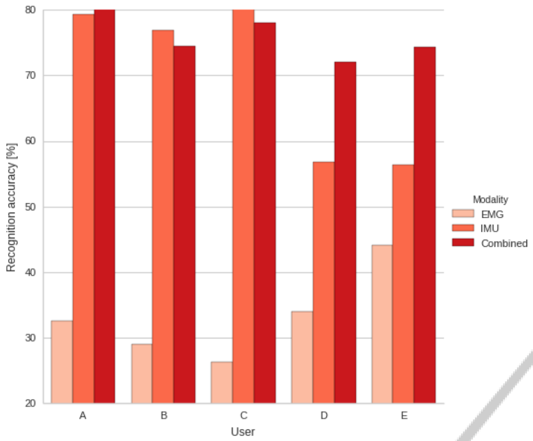
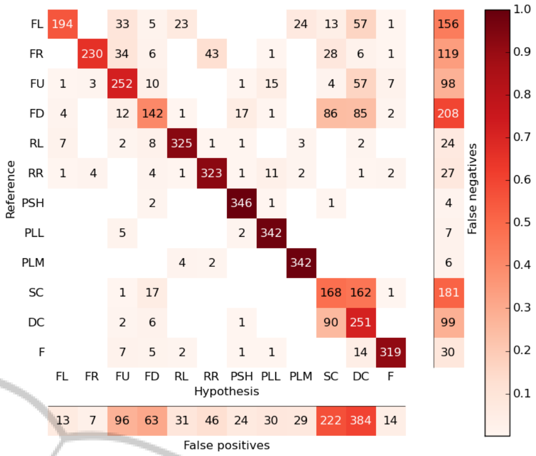


Figure 9: Person-independent recognition accuracy. Each bar depicts the achieved accuracy with the recordings of the respective subject used as test samples. For each subject, the performance when using IMU or EMG individually, as well as in combination, are shown.



	FL	FR	FU	FD	RL	RR	PSH	PLL	PLM	SC	DC	F	False negatives
FL	194		33	5	23				24	13	57	1	156
FR		230	34	6		43		1		28	6	1	119
FU	1	3	252	10			1	15		4	57	7	98
FD	4		12	142	1		17	1		86	85	2	208
RL	7		2	8	325	1	1		3		2		24
RR	1	4		4	1	323	1	11	2		1	2	27
PSH				2			346	1		1			4
PLL			5				2	342					7
PLM					4	2			342				6
SC			1	17						168	162	1	181
DC			2	6			1			90	251		99
F			7	5	2		1	1			14	319	30
False positives	13	7	96	63	31	46	24	30	29	222	384	14	

Figure 10: Confusion matrices for person-independent gesture recognition. The labels follow Table 2.

pate that the direction of a movement is largely encoded in the IMU signal, we expect that gestures are confused, that differ mostly in the direction of their movement. This is the case with *flick left*, *flick right*, *flick up* and *flick down*, all being performed mainly with extended index and middle finger. Also *flat hand push* and *pull*, as well as *palm pull* are sometimes confused. As the accuracy is still rather high, one can assume that the direction of movement does in fact imprint a pattern on the activity signals.

Making a fist produces very distinctive features in all feature space dimensions for EMG, as all the monitored muscles contract. Accordingly, in contrast to IMU based recognition, only few false positives and negatives are reported for the gesture *fist*.

4.2 Person Independent

For person-independent evaluation, the recordings of all subjects but one are used as training data in the cross-validation, leaving the recordings of the remaining subject as test data. The same model and preprocessing parameters as for session-independent testing are used.

We achieve a mean accuracy of 74.3%(±4.97%) for person independent recognition. The results for each test run, and therefore for each subject, are shown in Figure 9, together with the results when using the modalities individually.

Figure 10 shows the confusion matrix for person-independent recognition. As expected, the matrix does not show a diagonal character as pronounced as in the confusion matrices for session independent recognition. The gestures *flat hand push* and *pull*, *palm pull*, *fist* and *rotate left* and *right* are the classes with the lowest false negative count. The classes, together with the *flick left* and *right* classes, also have the lowest false positive count. Person-independent recognition seems therefore already reliable for a specific gesture subset. But especially the *flick* gestures, as well as the *click* gestures are often confused with others. Presumably, the inter-person variance is the highest for these classes.

4.2.1 Modality Comparison

Also for person-independent recognition the performance of the individual modalities is evaluated.

Figure 9 compares the best results for each individual subject as well as the average result for both modalities. Clearly the IMU with an average accuracy of 70.2%(±11.21%) outperforms the EMG modality with only 33.2%(±6.06%) accuracy on average. But the multimodal recognition still benefits from the combination of both sensor modalities.

Figure 11 shows the misclassifications when using only features extracted from the IMU signals. The confusion matrix is largely similar to the one resulting from a combined usage of both modalities. This underlines that for this system, EMG has rather small influence on person-independent recognition results. But some differences have to be addressed. The gesture *double click* shows a higher false positive count and is often given as the hypothesis for other gestures. As was mentioned before, we assume that especially the EMG signals are descriptive for *fist*. It it therefore consistent to this expectation, that the most prominent degradation of performance is visible for the *fist* gesture, which was performing rather well when using IMU and EMG together. Both its false positive and

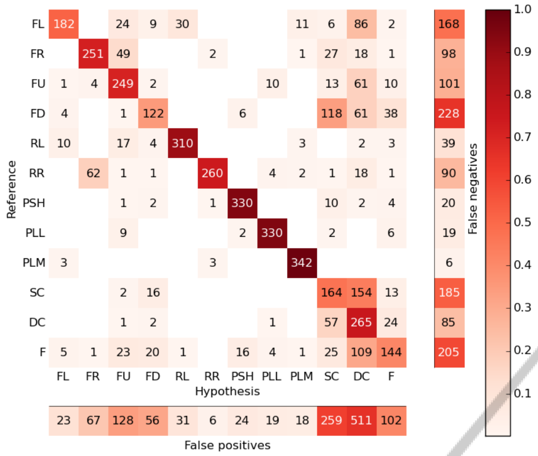


Figure 11: Confusion matrices for person-independent gesture recognition. The recognizer used only the features extracted from the IMU signals. The labels follow Table 2.

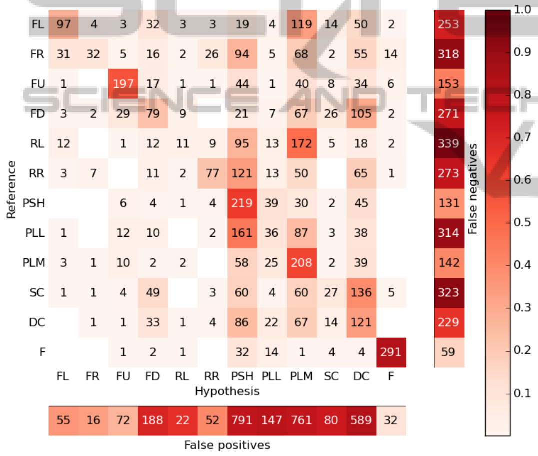


Figure 12: Confusion matrices for person-independent gesture recognition. The recognizer used only the features extracted from the EMG signals. The labels follow Table 2.

negative count are now rather high.

The diagonal character of the confusion matrix in Figure 12, representing recognition using only EMG features, is rather weak. Interestingly, no predominant attractor classes are present. Also no distinct block formations are visible, meaning that there is no pronounced clustering of different gesture groups in feature space. Instead, inter-person variance leads to many confusions scattered across the confusion matrix. The gestures *flick upwards* and *fist* both performe exceedingly well using only EMG, thereby explaining why they also perform well using both modalities combined.

We expected person-independent gesture recognition with EMG to be a hard task, as EMG signals vary from person to person and produce patterns over the different channels that are hard to compare. This explains, why only a few studies exist to person-independent, solely EMG based gesture recognition. Still, with 33.2% accuracy we perform better than chance with 8.33%.

5 CONCLUSION

In this paper we present a system for recognizing hand and finger gestures with IMU based motion and EMG based muscle activity sensing. We define a set of twelve gestures and record performances of these gestures by five subjects in 25 sessions total. The resulting data corpus will be made publicly available. We built a baseline recognition system using HMMs to evaluate the usability of such a system. We achieve a high accuracy of 97.8% for session-independent and an accuracy of 74.3% for person-independent recognition, which still has to be considered rather high for person-independent gesture recognition on twelve classes. With that, we show both the feasibility of using IMUs and EMG for gesture recognition, and the benefits from using them in combination. We also evaluate the contribution of the individual modalities, to discuss their individual strengths and weaknesses for our task.

ACKNOWLEDGEMENTS

This research was partly funded by Google through a Google Faculty Research Award.

REFERENCES

Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition Computation*, 5(4):307–331.

Amma, C., Georgi, M., and Schultz, T. (2014). Airwriting: a wearable handwriting recognition system. *Personal and Ubiquitous Computing*, 18(1):191–203.

Benbasat, A. Y. and Paradiso, J. A. (2001). An inertial measurement framework for gesture recognition and applications. In Wachsmuth, I. and Sowa, T., editors, *Revised Papers from the International Gesture Workshop on Gesture and Sign Languages in Human-Computer Interaction*, volume LNCS 2298 of *Lecture Notes in Computer Science*, pages 9–20. Springer-Verlag.

Chen, X., Zhang, X., Zhao, Z., Yang, J., Lantz, V., and Wang, K. (2007). Hand gesture recognition research based on surface emg sensors and 2d-accelerometers. In *Wearable Computers, 2007 11th IEEE International Symposium on*, pages 11–14.

Cho, S.-J., Oh, J.-K., Bang, W.-C., Chang, W., Choi, E., Jing, Y., Cho, J., and Kim, D.-Y. (2004). Magic wand: a hand-drawn gesture input device in 3-d space with inertial sensors. In *Frontiers in Handwriting Recognition, 2004. IWFHR-9 2004. Ninth International Workshop on*, pages 106–111.

Hartmann, B. and Link, N. (2010). Gesture recognition with inertial sensors and optimized dtw prototypes. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Istanbul, Turkey, 10-13 October 2010*, Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Istanbul, Turkey, 10-13 October 2010, pages 2102–2109. IEEE.

Hauptmann, A. G. and McAvinney, P. (1993). Gestures with speech for graphic manipulation. *International Journal of ManMachine Studies*, 38(2):231–249.

Kim, D., Hilliges, O., Izadi, S., Butler, A. D., Chen, J., Oikonomidis, I., and Olivier, P. (2012). Digits: Freehand 3d interactions anywhere using a wrist-worn gloveless sensor. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, UIST '12, pages 167–176, New York, NY, USA. ACM.

Kim, J., Mastnik, S., and André, E. (2008). Emg-based hand gesture recognition for realtime biosignal interfacing. In *Proceedings of the 13th International Conference on Intelligent User Interfaces*, volume 39 of *IUI '08*, pages 30–39, New York, NY, USA. ACM Press.

Li, Y., Chen, X., Tian, J., Zhang, X., Wang, K., and Yang, J. (2010). Automatic recognition of sign language subwords based on portable accelerometer and emg sensors. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*, ICMI-MLMI '10, pages 17–1, New York, NY, USA. ACM.

Mistry, P., Maes, P., and Chang, L. (2009). WUW-wear Ur world: a wearable gestural interface. *Proceedings of CHI 2009*, pages 4111–4116.

Oviatt, S. (1999). Ten myths of multimodal interaction. *Communications of the ACM*, 42(11):74–81.

Rabiner, L. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

Rekimoto, J. (2001). GestureWrist and GesturePad: unobtrusive wearable interaction devices. *Proceedings Fifth International Symposium on Wearable Computers*.

Samadani, A.-A. and Kulic, D. (2014). Hand gesture recognition based on surface electromyography. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*.

Saponas, T. S., Tan, D. S., Morris, D., and Balakrishnan, R. (2008). Demonstrating the feasibility of using forearm electromyography for muscle-computer interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 515–524, New York, New York, USA. ACM Press.

Wolf, M. T., Assad, C., Stoica, A., You, K., Jethani, H., Vernacchia, M. T., Fromm, J., and Iwashita, Y. (2013). Decoding static and dynamic arm and hand gestures from the jpl biosleeve. In *Aerospace Conference, 2013 IEEE*, pages 1–9.

Zhang, X., Chen, X., Li, Y., Lantz, V., Wang, K., and Yang, J. (2011). A framework for hand gesture recognition based on accelerometer and emg sensors. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 41(6):1064–1076.