

Device Control Using Gestures Sensed from EMG

Kevin R. Wheeler
NASA Ames Research Center
Moffett Field, CA 94035
E-mail: Kevin.R.Wheeler@nasa.gov

Abstract— In this paper we present neuro-electric interfaces for virtual device control. The examples presented rely upon sampling Electromyogram data from a participants forearm. This data is then fed into pattern recognition software that has been trained to distinguish gestures from a given gesture set. The pattern recognition software consists of hidden Markov models which are used to recognize the gestures as they are being performed in real-time. Two experiments were conducted to examine the feasibility of this interface technology. The first replicated a virtual joystick interface, and the second replicated a keyboard.

I. INTRODUCTION

We introduce an approach of designing and using neuro-electric interfaces for controlling virtual devices. Hand gestures are used to interface with a computer instead of manipulating mechanical devices such as joysticks and keyboards. Electromyographic signals (EMG) are non-invasively sensed from the muscles used to perform these gestures. These signals are then interpreted and translated into useful computer commands. We have chosen to demonstrate replacing joysticks and keyboards in order to present a platform with which the average computer user has had some experience. Ultimately, more evolved computer interfaces will not involve the movements associated with joysticks and keyboards but will use more inherently natural movements for expression. We envision that multiple modalities will be involved such as combining speech with gestures, and that these gestures will evolve to fit the most natural application.

There are a multitude of applications for bio-electric interfaces. Some of the applications that we are considering include the use of EMG to control a robotic arm or a robotic exoskeleton. In these cases the mechanical device could be made to be controlled as a natural extension of a person's body rather than requiring special training to work multi-degree of freedom joysticks. A space-based application could include the capability of typing into a computer while being restricted by a space suit. In the event that there is a depressurization accident on a long term space mission, and the astronauts need to access on-board computers, EMG electrodes within the space suit could be used

to replicate a computer interface. Another application includes the control of remote vehicles. Examples could include mobile search and rescue robots and unmanned aircraft (UAVs). In the case of UAVs we envision a "wearable cockpit". This concept involves a pilot wearing computer display goggles which would display the cockpit instrumentation. The pilot could then interact with the virtual switches, dials, and control sticks by performing the required gestures. This would allow for easier reconfiguration of mission instrumentation providing the ability to have the instrumentation change during the mission and also allowing each pilot to have his own preferences reflected in the instrumentation.

Currently most gesture recognition systems come in one of two forms:

- Gestures are recognized via an external camera which requires sophisticated image processing and controlled lighting.
- Gestures are recognized by placing a sensing glove on the hand(s) of the participant.

We desire to do recognition in poor lighting conditions in extreme environments (outside of the lab) without encumbering our hands and without invasive procedures. To date, we have accomplished this by directly connecting a person to the computer via EMG surface electrodes on the forearm. The EMG signals are sampled, digitized, and the resulting time-series are passed through a pattern recognition system based upon hidden Markov models (HMMs). The recognized patterns are transmitted as computer commands. Our first example of this ([1]) was to attach four pairs of electrodes to one forearm, and then interpret the resulting EMG signals as joystick commands. These commands were then used to fly a realistic flight simulator for a 757 transport aircraft. The acting pilot would reach out into the air, grab an imaginary joystick, and then pretend to manipulate this stick to achieve left and right banks and up and down pitches of the aircraft simulation.

Our latest results are in pretending to type on a table (or lap) and translating the resulting sensed EMG signals into

keystrokes. In the following sections we will describe the methods we used to acquire the EMG data, to recognize the patterns and the experimental tasks that we controlled with these signals. Lastly, the results and a conclusion will follow.

II. METHODOLOGY

Each gesture set type necessitated a different methodology be used. The virtual joystick gesture set used four pairs of dry electrodes and four coarse grained movements. The virtual keyboard gesture set consisted of 8 pairs of wet electrodes and 11 fine grained movements. The methodology that we followed consisted of the following steps:

1. Gesture selection
2. Electrode application (location and number)
3. Signal acquisition, filtering, and digitization
4. Feature formation
5. Pattern recognition model training and testing
6. Pattern recognition application in interactive simulation

Each of these steps will now be described in detail. The process started by selecting the desired physical motions (gestures) to be used to control the virtual device. From the set of gestures, the best location for the limited number of electrode pairs (a maximum of 8 in our case) was established. Then standard signal processing practices were used to filter and digitize the signal. Transforms such as moving averages were applied to this raw digital data. The transformed data was fed into the pattern recognition software to train the models. Once the pattern recognition models were trained, they could be used for the real-time recognition task.

A. Gesture selection

Our first task used coarse grained gestures to mimic manipulating a joystick ([1]). There were four basic gestures pretending to move a joystick: up, down, left, and right. The recognition of these gestures was made easier because there were only four gestures and there were four pairs of electrodes which provided for reasonable separation between gestures.

Our second task consisted of pretending to type on a number pad on the keys 0 - 9 and Enter. These movements consisted of much finer grained gestures. The first, second, and third fingers were resting over the 4, 5, and 6 keys respectively. The first finger was used to press the keys 1, 4, and 7. The second finger was used to press the keys 2, 5, and 8. The third finger struck the keys 3, 6, and 9. The fourth finger was used for the Enter key. The thumb was used to strike the zero key. In this case we used 8 pairs of electrodes.

B. Electrode application

The location of the electrodes depend upon the gestures that we wish to recognize and upon individual physiological differences. We speculate that it is possible to have gesture independent placement if enough electrodes are

used and then an automated sub-sampling procedure initiated. A practical drawback to this technique is that over 50 channels are required to be simultaneously sampled to fully cover the forearm. Another problem with gesture independent placement is that the active channels will not in general be applicable to new gestures not yet added to the activation set. In the future, this problem will be obviated with smaller sensors which can be spread onto the skin by the hundreds with corresponding shrinkage in the pre-amplifiers.

The joystick task required 4 electrode pairs. Dry electrodes sewn into a sleeve were used as shown in Figure 1. This sleeve helped to reduce variation in the placement of the electrodes. The typing task required 8 electrode pairs. We chose to use wet electrodes for this work due to the improved signal to noise characteristics of wet electrodes over that of dry electrodes. This was in part due to the signal amplitudes for the typing task being much smaller than that of the joystick task. The drawback to using wet electrodes is that the position of the electrodes unintentionally varied from one day to the next. The locations of these pairs were obtained by establishing a grid of electrodes on the forearm, and then performing the desired task; and only those electrodes which produced distinct signals for a gestures were used. The positions of the electrodes for the typing task were in two rings around the forearm: one near the wrist, and one near the elbow also shown in Figure 1.

Several tests were conducted to measure the effects of minor variations in placement (1-3 mm) and major displacements (1-2 cm). The minor variations had no impact but the major displacements require that the recognition models be re-trained or adapted for the individual user. Individual differences in personal physiology proved to be challenging. Differences in arm lengths and widths made it difficult to place the electrodes at the proper positions across people without considerable effort. In addition, strengths of the EMG signals varied across people and varied with the amount of training that individuals received.

C. Signal acquisition, filtering, and digitizing

The EMG data was acquired by placing differential instrumentation pre-amplifiers near to each electrode pair with a Common Mode Rejection Ratio of 110 dB. All eight channel pairs were referenced from a common ground electrode positioned over bone at the wrist or elbow. The signal was digitized using 16 bits at 6000 Hz., and then a 32 tap anti-alias bandpass Bessel filter was applied and then down sampled to 2000 Hz unless otherwise indicated.

D. Feature formation

The goal of the feature formation step is to separate the signals enough to allow the pattern recognition module to distinguish between gestures. Another result of the transformation is to create a space smooth enough to be reliably modeled. We tried many common methods such as Short Time Fourier Transform (STFT), wavelets, moving aver-

ages, and auto-regression coefficients. In the end, the simplest feature space seemed to be the best: moving averages. Since the EMG signals were differentially amplified, the average of the signals when presented with enough samples was approximately zero. This required that the moving average be performed on the absolute value of the signals. The windows used to form the moving averages were allowed to overlap by 75 percent. Note that this is purely an amplitude-based method; the frequency of the electrical activity did not seem to vary significantly from one gesture to the next.

E. Pattern recognition

The pattern recognition method we chose to employ was a hidden Markov model (HMM). HMMs have been developed by the speech recognition community in response to their pattern recognition time-series problem ([2]). The history of speech recognition reveals a process which first attempted to recognize isolated words from a single speaker, then isolated words from multiple speakers, followed by continuous words from a single speaker, and finally continuous words from multiple speakers. We are following a similar approach with our gesture recognition work. We have developed isolated gesture recognition for both a single participant and for multiple participants. The work described in this paper will describe isolated recognition for a single typist and continuous recognition for the joystick study.

Two issues with training any model to learn from sampled data are that the dataset is representative and that the model has the appropriate number of parameters for accurate representation. The training dataset can suffer from not having enough exemplars or it is inconsistent for the sample size. In our case, we can always sample more data if we do not have enough. On an empirical basis, we have experienced being able to use as few as 20 exemplars from each gesture to adequately model the remaining data from a single day. However, when we combine data from multiple days it becomes readily apparent that inconsistency is a problem.

We define inconsistency as the statistics of the data varying from day to day. We could have defined this in terms of gesture to gesture variation but have chosen not to because this variation is more of a natural variation inherent in human behavior whereas the day-to-day inconsistencies are more an artifact of the experimental procedures.

There are many solutions to resolving this inconsistency. There are many contributions to the variations which could be minimized. One example is electrode placement. If the electrode locations are allowed to vary from day to day then the signal statistics will also vary. This can be eliminated through the use of a fixed electrode sleeve.

Day to day variations related to natural behavior may not be possible to remove, and in fact we would benefit from modeling them. One example is the way that people gesture may vary slightly from day to day even though their intentions are to perform the gestures identically from one

day to the next. In this case we need to have enough data to represent the multi-modal statistics and we need a way to adapt the system models from day to day. Our current methodology does not vary adaptively but it is our plan to include this in future work. This means that our best remedy is to recognize when day to day variation is too great for adequate model generalization. We can then use less data for training by only using the data similar to our current day's setup (i.e. electrode locations).

E.1 training

The HMMs we used were continuous, tied mixture ([3], left to right models. Standard Baum-Welch training ([2]) was used. For those readers not familiar with HMMs, here is a brief explanation of each of these terms. The models are called *continuous* if they use inputs which can take on a range of floating point values. The alternative to this is to allow for only discrete values such as might be found if the input were transformed by quantization. *Tied mixtures* means that a fixed number of Gaussian mixtures are used through out all of the states. Thus any state may make use of any mixture. A *left to right* model means that the HMM may not go back to a previous state but may remain in a state or go onto a new state.

The initialization of the models were performed using K-means clustering. The states were partitioned to equalize the amount of variance present within each state. The data sets used to train were segmented to insure that the peak of the variance was near the middle of each segment. This translated to the bulk of the energy being centered. Segments were sampled at 2000 Hz and contained 3072 samples per channel, with eight channels total. The parameters of the HMMs that we typically varied were the number of discrete states, the number of Gaussian mixtures, the number of maximum number of iterations to train, the method used to arrive at the state partitioning (uniform vs. variance based), and the method used to initialize the parameters of the mixtures (e.g. k-means clustering).

E.2 recall

The real-time recall was performed using the standard Viterbi algorithm ([4]). Since the system was processing data as the data streamed by, there was no knowledge as to where the peak of the variance was occurring. Because of this, the HMMs would see the data when the peak was first at the left most in the time segment, then the peak would move across from left to right, and then the final presentation was when the peak was at the right most part of the segment. Since the HMMs were trained only when the peak was centered, due to this shifting, the HMMs were required to recognize a gesture several times in a row before that gesture was selected as the one that was observed. This prevented spurious recognition when the peak was not near the center of observation.

F. Experiments

The experiments were conducted to determine the feasibility of achieving two goals:

1. Substituting for a joystick with bioelectric signals
2. Substituting for a keyboard with bioelectric signals

The first experiment consisted of four pairs of dry electrodes fitted within a sleeve to the forearm of a participant. The participant was then asked to pretend to be moving a joystick left, right, up, and down. The participant performed each of these gestures 50 times. The data was separated by gesture, and segmented to have the peaks be in the center of 3072 sample segments. Artifacts or incomplete gestures were removed from the data sets via manual inspection. These sets were then used to train four HMMs, one for each gesture. These trained models were then used to recognize gestures made on a day excluded from the training set using the same dry electrode sleeve. A confusion matrix was generated to display errors and to show which gestures were confused with one another. The system has also been used for numerous real-time demonstrations of flying a simulated 757 transport aircraft to landing ([1]). A more continuous gesture recognition was implemented by decreasing the segment size.

Four methods were used to test the pattern recognition system. The first involved training the models on data from one day, and then recalling on different data obtained on the same day. We will call this method same trial acquisition and testing. The second involved training on data from one day and recalling on data collected on a different day. We will call this method cross-trial acquisition and testing. The third method trained on data sub-sampled from a large set taken across multiple days, and then recall was performed on data different from the training but in the same large set. This third method will be called multi-trial acquisition and testing. The final method involved training on a previously acquired single day that provided the best recognition in our real-time simulation for flying an aircraft. We will call this best trial training and real-time testing.

The second experiment used 8 pairs of wet electrodes in two rings of four each, one ring near the wrist, and the second near the elbow. The participant was asked to touch type on a printed picture of a number pad keyboard, striking the keys 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 and Enter. The participant was asked to type these in order, separated by a one second rest interval, for a total of 40 strokes on each key. This data was then segmented, and artifacts were manually removed. Data was collected on several different days. Eleven HMMs were trained, one for each gesture.

The performance on batch data sets is not equivalent to the performance found in live demonstrations. Typically batch data sets are collected under static conditions. The live demonstrations are typically performed under high stress, with imperfect electrode placement, while the participant is bombarded with questions and distractions. Thus live performance tends to suffer from more

TABLE I
CONFUSION MATRIX FOR CROSS-TRIAL JOYSTICK DATA

Gesture	Left	Right	Up	Down	Correct
Left	15	0	26	9	30%
Right	0	50	0	0	100%
Up	0	0	50	0	100%
Down	0	0	1	49	98%

errors than the batch testing results.

III. RESULTS

A. Joystick

A.1 Same trial acquisition and testing

This experiment is by far the easiest to recognize because the variability associated with day-to-day differences has been eliminated. Such variation includes conductivity levels of the skin, positioning of the dry electrode sleeve, and changes in the performance of the gestures. We noticed that we could determine when participants had used skin moisturizer before the experiment because the signal quality obtained from the dry electrodes improved. Typically if the HMMs had an appropriate number of parameters and enough data was used, then no errors were made upon recall of the validation set.

A.2 Cross-trial acquisition and testing

This experiment demonstrated which gesture was the hardest to separate from the others. In particular, we trained on 50 instantiations from one trial date and then validated on 50 other instantiations from a different trial date. A typical confusion matrix is shown in Table I. This indicates that day-to-day variations were significant enough to cause difficulty in separating the gesture Left from other gestures. The source of the variations included electrode placement, length of the gesture, strength of gesture formation, and the form of the gesture (wrist angles). This lead to the next experiment to see if the models would generalize if we trained on all of the different days together and then tested on a withheld subset.

A.3 Multi-trial acquisition and testing

To determine the generalization capability of the HMMs we trained on data from multiple trial dates and then recalled on points withheld from the training data for the same dates. This resulted in good results indicating that the models had sufficient number of parameters to learn the data complexities (although arguments for over-fitting could be made, the real-time recall results are what is most important). A typical confusion matrix is shown in Table II.

Of course this does not mean that when yet another new day of data is added that the system will be able to generalize on that data. In the next experiment we examine

TABLE II
CONFUSION MATRIX FOR MULTI-TRIAL JOYSTICK DATA

Gesture	Left	Right	Up	Down	Correct
Left	50	0	0	0	100%
Right	0	50	0	0	100%
Up	0	0	50	0	100%
Down	0	0	0	50	100%

TABLE III
CONFUSION MATRIX FOR CROSS-TRIAL SHORT JOYSTICK DATA

Gesture	Left	Right	Up	Down	Correct
Left	17	0	25	8	34%
Right	0	49	0	1	98%
Up	0	0	50	0	100%
Down	0	0	0	50	100%

training on the single best day and using that for real-time testing. The real-time testing acts as new and unseen data.

A.4 Best-trial training and real-time testing

The error rates determined from the previous methods were not necessarily indicative of real-time performance. In particular the error rates would vary across time depending upon many factors such as sleeve position (rotation), sweating, skin moisture (dry skin does not conduct well), length of time that the electrodes were worn, and fatigue (resulting in tremors). By training on only a single day's data, we were able to use the dry electrode sleeve for demonstrations on many different days. We selected the day which gave the best real-time reliability. The demonstrations consisted of flying a 757 transport aircraft to landing at a simulation of San Francisco airport.

A.5 Continuous Recognition

In the previous experiments, a total of 3072 data samples were used to form the estimate. This introduced considerable time lag into the system (1.5 seconds). In an attempt to become closer to a continuous recognition process, the time segments used to train the HMMs were shortened to only contain the first part of the rise of the signal using 352 samples. In this case the HMMs consisted of 3 states with 9 mixtures total. The resulting cross-trial confusion matrix is shown in Table III.

This matrix is not significantly different from the previous cross-trial longer data. The change in signal length allowed for us to remove noticeable delays between the gesture action and the movement of the aircraft. We now have less noticeable delays (176 ms.) but this in turn has caused a slight decrease in the robustness of the gesture recognition process. It is possible to halve this response time with a small decrease in the recognition rates. Since this is intended for real-time systems, such a lag is hard to justify but it has not prevented us from successfully flying the simulated aircraft. The resulting multi-day confusion

TABLE IV
MULTI-TRIAL CONFUSION MATRIX FOR TYPING DATA

	1	2	3	4	5	6	7	8	9	%
1	46	0	0	4	0	0	1	0	0	90
2	0	48	0	0	0	0	0	3	0	94
3	0	0	49	0	0	1	1	0	0	96
4	11	0	0	38	2	0	0	0	0	75
5	1	3	0	5	36	1	3	2	0	71
6	0	1	6	0	0	42	0	0	2	82
7	0	0	0	0	0	0	51	0	0	100
8	0	0	0	0	2	1	3	44	1	86
9	0	0	0	0	0	0	0	0	51	100

matrix had no errors. In the next set of experiments we have switched from using the dry electrode sleeve to wet electrodes. This was necessary because the EMG signals measured for the typing gestures were much smaller than those for the joystick. The wet electrodes tend to have a higher signal-to-noise ratio than the dry electrodes.

B. Keyboard

In this experiment, the position of the hand above the simulated number pad was maintained in a touch-typist typing position. If the position of the hand were allowed to vary, the tasks of distinguishing between hitting the top row of keys from the bottom row of keys would greatly increase in difficulty and would require electrodes on the upper arm to sense the movement. The angle of the participant's wrist also had to be carefully maintained to avoid radically changing the sampled signals. Even with careful attention to position and maintaining electrode placement from day to day, the data tended to vary. There are consistent trends in the data but that the variation between instantiations of the same gesture is great. The impact of these variations will be discussed further in the results section.

B.1 Multi-trial acquisition and testing

In the keyboard replication experiments we had much greater daily variation in electrode placement than with the joystick. We also had difficulty in reliably having the participants maintain a consistent hand position from trial to trial. This included making sure that the wrist angle was similar and that the hand was consistently either not resting on the table during motion or was in part supported by the table (i.e. bad form, but consistent bad form). Given all of these difficulties, the resulting confusion matrix shown in Table IV for multiple trials looks pretty good. The variability in the data caused our models to generalize gestures such that more confusion occurred. For live demonstrations, we needed to train on the same day that we were giving the demonstration, and thus only used a single day's data. However, the training only took a few seconds.

The use of wet electrodes has caused unintentional misplacement. We are currently developing new dry electrode straps which have a higher density than the sleeve and

which are similar in size to a wrist band. These straps will allow for us to have the electrodes on a band positioned relative to each other without variability. The second enhancement is to include model correcting adaptation which is now common in the speech recognition community. This adaptation would allow for the models to be tuned to small variations both throughout the day and also if there are differences between the models used to train and the current day's configuration. It will also be possible to have a calibration stage such that the participant can make a gesture to issue a certain command and the computer will adapt to understand the signals as that command. This will allow for us to get away from requiring a participant to learn a fixed set of gestures. The person will be able to perform a gesture that seems natural to him or her in order to accomplish a given task, and the computer will simply map those signals to the correct action.

IV. CONCLUSION

We have shown that it is feasible to control virtual devices via non-invasive EMG signal monitoring. We have demonstrated that it is possible to replicate both joysticks and keyboards. We chose to demonstrate these input devices because they are familiar to the general computer user. Ultimately, the most natural interface will consist of more natural movements than those associated with either joysticks or keyboards. When our on-line adaptation software has been completed it will be possible for a person to make the gesture he feels is natural for a given task (assuming that we have enough electrodes to cover the muscles involved) and the computer will have to map from this "natural" signal space to the expected "computer command" space.

Although the dry electrode sleeve guaranteed that the electrodes would be positioned consistently relative to each other, there was no guarantee that the sleeve would be in the exact same location on the arm. It was our hope that the sleeve would minimize day to day variations, but we need to redesign the sleeve to assure positioning. We would also like to increase the number of electrodes to allow for spatial oversampling. The dry electrode sleeve suffers from intermittent conductivity problems. Sometimes the impedance between a dry electrode and the skin will become temporarily elevated. This sometimes happens due to hairs lifting the sensor or variations in moisture levels in the skin. This can usually be identified in the signals and the user subsequently notified that the electrode needs to be wiggled to re-establish contact. This is, of course, not acceptable for daily use.

Ultimately, we see that there are a variety of applications for this work. The ability to naturally interface with a computer allows for humans to manipulate any electrically controlled mechanical system. In addition to wearable computing applications we are also examining interfaces to robotic arms, mobile robots for urban rescue, unmanned aircraft drones, robotic exoskeletons, and space suit interfaces. There are also side benefits to using EMG

signals for control in long duration space missions. One of the side effects of living in a zero gravity environment for extended periods is muscle atrophy. It would be possible to have astronauts train during a long flight to a distant planet by simulating the motions necessary to accomplish a given task. The EMG signals generated from these motions could be analyzed. If significant variations were detected, the astronaut could be given advanced warning to change their training routines to ensure mission success.

Our next generation work is to integrate both EEG signals and EMG signals so that we have an augmented brain-computer interface. Since EMG signals are much faster than EEG signals, this will allow us to have both a fast response EMG interface and a more slower response thought based interface.

REFERENCES

- [1] Charles Jorgensen, Kevin Wheeler, and Slawomir Stepniewski. Bio-electric flight control of a 757 class high fidelity aircraft simulator. In *Proceedings of the World Automation Congress*, 2000.
- [2] Lawrence R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257 – 286, February 1989.
- [3] Jerome R. Bellegarda and David Nahamoo. Tied mixture continuous parameter modeling for speech recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(12):2033 – 2045, December 1990.
- [4] G. David Forney Jr. The viterbi algorithm. *Proceedings of the IEEE*, 61(3):268 – 278, March 1973.

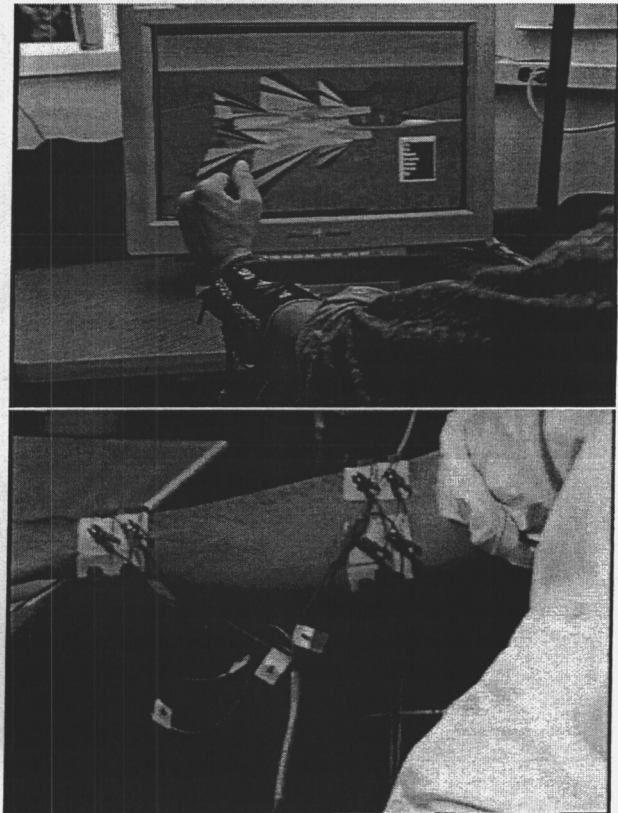


Fig. 1. Dry Electrode sleeve for joystick based flying, and wet electrodes for typing experiments