Conversion of Sign Language to Spoken Sentences by Means of a Sensory Glove

Pietro Cavallo
Electronic Engineering Department, University of Tor Vergata, Roma, Italy
Email:p.cavallo85@gmail.com

Giovanni Saggio
Electronic Engineering Department, University of Tor Vergata, Roma, Italy
Email: saggio@uniroma2.it

Abstract—**Normal communication of deaf people in ordinary life still remains an unrealized task, despite the fact that Sign Language Recognition (SLR) made a big improvement in recent years. We want here to address this problem proposing a portable and low cost system, which demonstrated to be effective for translating gestures into written or spoken sentences. This system relies on a home-made sensory glove, used to measure the hand gestures, and on Wavelet Analysis (WA) and a Support Vector Machine (SVM) to classify the hand's movements. In particular we devoted our efforts to translating the Italian Sign Language (LIS, Linguaggio Italiano dei Segni), applying WA for feature extractions and SVM for the classification of one hundred different dynamic gestures. The proposed system is light, not intrusive or obtrusive, to be easily utilized by deaf people in everyday life, and it has demonstrated valid results in terms of signs/words conversion.**

Index Terms—**Machine intelligence, Pattern analysis, Human Computer Interaction, Support Vector Machines, Data glove, Sign Language Recognition, Italian Sign Language, LIS**

I. INTRODUCTION

S imilarly to spoken languages, Sign Languages (SLs) are complete and powerful forms of communication, and are adopted by millions of people, who suffer from deafness, all over the world. SLs are different among different regions and states: American Sign Language (ASL), Japanese Sign Language (JSL), German Sign Language (GSL), Lingua Italiana dei Segni (LIS, the Italian Sign Language), etc. However, each single SL relies commonly on gesture and posture mimic interpretation, which plays a fundamental role in non-verbal communication too. SL comprehension is generally limited only to a restricted part of population, thus deaf people remain restrained apart from social interactions with hearing persons, and the body-language/non-verbal communication is mainly limited to "feelings" rather than consciousness

understanding. These are the main reasons why a system for "Automatic" Sign Language Recognition (A-SLR) is welcome and a greater human effort is being devoted to realize it [1]. A-SLR could allow deaf people to communicate without limitations, could assign suitable interpretations to non-verbal communication without ambiguities, and could be the basis of a new form of human-computer interaction, since the most natural way of human-computer interaction would be through speech and gestures, rather than the current adopted interfaces like keyboard and mouse. Particularly, the integration of A-SLR with automatic text writing or speech synthesis modules can furnish a "speaking aid" to deaf people.

Our purpose is to realize a system capable to measure human static and dynamic postures, classify them as "words" organized into "sentences", in order to "translate" SLs into written or spoken languages. To this aim, a great challenge comes from the measure of human postures with acquisition devices that are both comfortable and easy to use. In particular, we have to focus our attention on hand postures and movements since the SL is mostly, even if not exclusively, based on them. In fact SL is made of hand gestures, body and facial expressions, but the latter is not strictly "fundamental".

Currently, hand movements are commonly measured by motion tracking techniques based on digital cameras. These systems offer interesting results in terms of accuracy, but suffer from a small active range and disadvantages related to portability. In order to overcome these problems, new measuring systems have been developed, in particular the ones based on sensory gloves (i.e. gloves equipped with sensors capable to convert hand movements into electrical signals).

The first sensory gloves on the market [2-3] were obtrusive for movements, uncomfortable and capable to measure only a very low number of Degrees Of Freedom (DOF) of the human hand. Nowadays, commercial sensory gloves are quite light, comfortable and capable to measure up to 22 DOFs [4], covering flex-extension and abdu-adduction movements of the fingers and spatial arrangement of the wrist. However, the cost remains generally too high (tens of thousands of dollars) to be widely applied in everyday scenarios. Thus, new sensory gloves have been created by research groups all over the

world [5], proposing different raw fabric materials (stretchable, washable, light, comfortable to don), different kind of sensors to convert postures into electrical signals (based on optics, magnetic fields, inertia principles, etc.), different front-end electronic solutions, and so on.

Here, we designed and realized a sensory glove based on both bend and inertial sensors sewn onto a Lycra support. The overall system is explained in details in section IV.A.

The electric values coming from the sensors are processed by an ad-hoc designed electronic circuitry that is connected to a personal computer where digital data are on-line processed to obtain a classification of the measured gestures.

By means of the glove and the hardware, we measured the static and dynamic postures of the hands of two persons who performed the signs belonging to the LIS (Linguaggio Italiano dei Segni, i.e. the Italian sign language). Thanks to feature extraction and gesture classification, we "translated" in real-time a stream of LIS signs into Italian words and sentences, which can be converted into speech via commercially available synthesizer software.

Apart from sign language translation, our data glove could have a great impact on Human-Computer Interaction applications, being the user able to translate its actions and gestures into computer commands.

This is an interesting application, especially with new devices coming up in the next future (e.g. smart glasses) which need a natural way of interaction.

It also has great potential for video games and for all those scenarios where nonverbal communication can improve the way people interact with machines.

We are strongly convinced that a glove system is much more user friendly and natural to use, especially when interfaced to a smartphone via wireless/Bluetooth connection, while a camera-based system suffers indeed of a restricted visual field, and it is not portable in everyday life.

A-SLR can be considered close to speech recognition algorithms, even though speech recognition only deals with one signal while A-SLR has to handle multiple signals, i.e. hand shape, orientation, position and movement. In addition, the commercial inertial sensors (as the ones we used for the glove) have in general "noisy" problems that must be taken into account.

For the "recognition" of the hand's kinematic, we adopted wavelet analysis: the dynamic components of the gestures were described by wavelet coefficients and then furnished to a Support Vector Machine (SVM) module for the classification process. Three different chunking methods were first considered and then implemented in order to truncate the signs contained in a continuous sentence.

The remainder of this paper is organized as follows. Section II reviews the state of the art. In Section III there is a brief introduction to the Italian Sign Language. In Section IV, details of our system are given: the sensory glove (subsection A); the acquisition data circuitry (subsection B); proposition of three different chunking methods (subsection C); feature extraction and gesture classification (subsections D and E). In section V the experimental results are given. The conclusions are drawn in the last section.

II. RELATED WORKS

The most commonly adopted method to measure human movements relies on "motion tracking" procedures, which usually involves optical techniques (e.g. Optotrak Certus @ www.ndigital.com, OptiTrack @ www.naturalpoint.com/optitrack). E. J. Muybridge was a pioneer of optical analysis of movements, by investigating about human and animal movements through photograph sequences [6]. Even though the method at the beginning was inevitably inaccurate, nowadays the optical system has reached complete maturity, but with the drawbacks of high costs, the need to arrange a scenario with cameras, and the need for a high computational workload. As time passed, other methods were invented such as magnetic based tracking systems (see the 3D Space Fastrak as an example, www.polhemus.com), hybrid systems combining inertial sensors with an optical marking system (MoCap), or mechanical based system consisting of an exoskeleton made of lightweight aluminum rods that follow the motion of the user's bones (see the Gypsy 7 Torso Motion Capture System, www.metamotion.com). Finally, we must mention wearable sensor systems consisting of sensors located on the human body as the sensory gloves.

These gloves are quite precise to measure finger joints angles with a resolution of the order of one degree, but with the drawback of the imperfect space tracking of the hand, i.e. the location of the hand in the 3D surroundings. In previous works [7-10] the most adopted way to carry out the tracking was to use magnetic field devices such as the well known Polhemus Fastrak [11]. Even though it provides accurate information about trajectory and orientation of the object to be tracked, the cost is fairly high. In addition, it has a limited action range and it is not portable at all.

In [12] and [13] space tracking was carried out by using accelerometers. These kinds of inertial devices require little energy to work, are comparatively cheap and small enough to be integrated in a glove. Unfortunately, for the Fastrak previously discussed, the signals obtained from the accelerometers are more difficult to process and with lower spatial resolution.

According to our knowledge, only few works are related to inertial devices in A-SLR applications; in [12], they are used to track hand movements to recognize a set of 3 gestures; in [13] the authors carried out the classification of 17 different trajectories (not related to any sign languages), using fusion features based on Wavelet analysis to represent the trajectories. We here extend the work of [13], using Wavelet coefficients as well to represent the dynamic components of the gestures, but analyzing a larger set of 100 gestures belonging to a real sign language (i.e. LIS), and using an innovative low-cost acquisition device.

Regarding the classification part, in literature different

classification techniques have been used to recognize static or dynamic gestures [14]. The first A-SLR attempts were focused on static gestures (e.g. finger alphabets), and the recognition algorithms were not so similar to the ones used for isolated and/or continuous signs belonging to dynamic gestures. Since we are going to study the latter ones we decided to focus on previous related works on dynamic gesture recognition and to leave out the works on static gestures.

[7] is one of the first attempts made in A-SLR; it used 5 neural networks to recognize 200 different signs with a data glove equipped with Fastrak. In [15] the authors adopted an algorithm based on a feed forward Neural Network trained with a Backpropagation algorithm. In [16] a simple recurrent network was used to separate the signs in a continuous stream and a Hidden Markov Model (HMM) was trained to classify the gestures. In [8] temporal clustering and a variant of HMM, known as transition movement model, were combined together to classify a large set of Chinese signs, but still using a Fastrak device to trace the space path.

Very few works are focused on the Italian sign language (LIS) recognition. In [10] a Self-Organizing Map Neural Network was adopted and combined with lips movement recognition via optical techniques to chunk the words in a continuous stream. On the other hand, in [17] the authors realized a virtual avatar capable to move according to the signs of the Italian LIS.

III. ITALIAN SIGN LANGUAGE

Even though it is not legally recognized, LIS is a complex language taken as a standard way of communication among the Italian deaf community. As in the example of figure 1, each sign is identified by four parameters:

a. The location where the sign is made
b. The 'configuration', which is the shape assumed by the hand while performing the sign
c. The orientation of the hand's palm
d. The movement of the hand in the space.

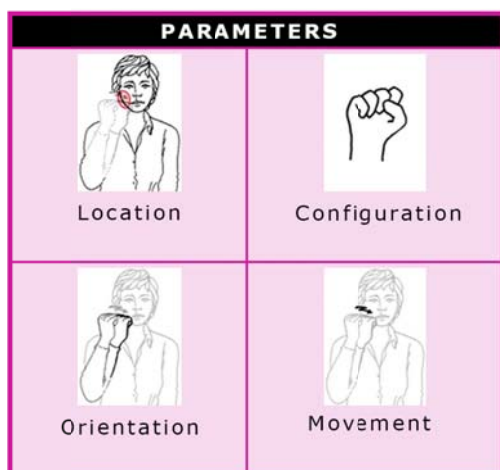For the purpose of this work, we limited our study to a

Fig. 1. Sign Language parameters

LIS subset of 100 signs taken randomly from the Italian Sign Language dictionary [18]. The number of signs here chosen is comparable to the average number found in literature.

IV. SYSTEM DESCRIPTION

Figure 2 depicts the general diagram of the system architecture.
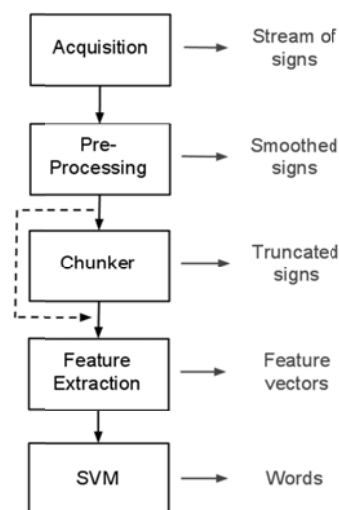

Fig. 2. General system architecture

Each block of the diagram, which represents the logical step of the overall method, has been realized by our research group named HITEG (Health Involved Technical Engineering Group).Wearing the data glove, the user is asked to perform a single gesture or a sequence of continuous signs. These are measured and signals are sent to a PC in a digital format. Data are then preprocessed to remove potential noise by means of a smoothing algorithm. If a sentence constituted by several gestures is recorded the Chunker will split it in single signs. We associate to each gesture a feature vector, which is then passed to the classifier, which utilizes it for learning or classification purpose.

A detailed explanation of all the logical steps can be found in the following paragraphs.

A. The Hiteg Data Glove

In order to measure hand movements we used our homemade sensory glove, shown in figure 3.

It is made out of a Lycra support and equipped with 15 bend sensors, to measure the movements of each distal interphalangeal joint, proximal interphalangeal joint and metacarpophalangeal joint. The resistance value of each bend sensor varies according to the amount of bending, so we can measure flexion and extension movements. No sensors were devoted to abduction and adduction movements since their influence on the measurement of the postures is negligible to the aim of our work. Moreover, to provide information about the space position of the hand, two inertial measurement units (IMU) Analog Combo Board Razor from SparkFun,

equipped with a 3 axis accelerometer and a gyroscope, are integrated in the glove, one in correspondence of the back of the hand and the other right below the elbow. The IMU sensor on the elbow is attached with a removable armband, and it will soon become wireless in order to make the user experience more comfortable.

Fig. 3. The HITEG sensory glove

B. Data Acquisition

In figure 4, the hardware needed for data acquisition is shown. Resistance signals from bend sensors are acquired at a sampling rate of 720 Hz per sensor, and converted into voltage units in a 10-bit digital format, by means of a voltage divider, a microcontroller and an ADC unit, so resulting 1023 logical levels ($0 \div 2^{10}$-1), matching an interval of $0 \div 4.99$V analogical values.
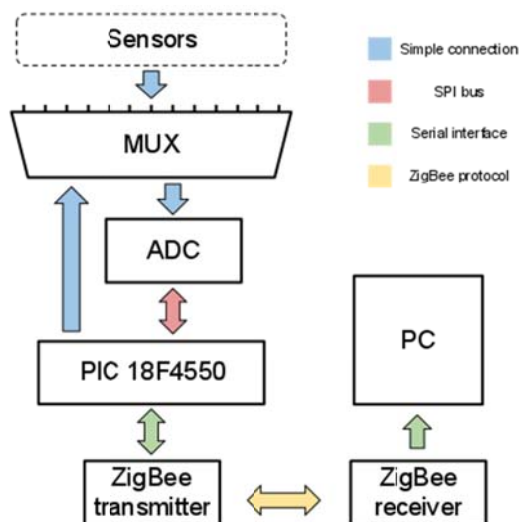

Fig. 4. Hardware architecture for data acquisition.

Both IMUs provide three values (in the range of $\pm 3g$) associated to the acceleration on the three axis, and three values for the angular speed (in the range of $\pm 300°/sec$) along each axis in the 3D space. Data are acquired with a PIC 18F4550 microcontroller, digitalized and provided to a computer via ZigBee protocol. Because of possible noise, the acquired data are then filtered using a moving average

smoothing (eq. 1), so that unwanted high frequency components (e.g. noise spikes) are discarded.

Given a time series $\{y_t\}$, ma is the moving average of order $k = m_1 + m_2 + 1$:

$$ma = \sum_{i=-m_1}^{m_2} \theta_i\, y_{t+i} \qquad (1)$$

where θ_i is the weight associated to the i-th observed value.

C. Chunking

To obtain a correct A-SLR, the recognition of each sign in a continuous stream (that constitutes a sentence) plays a fundamental role. The solution is far from trivial, also because each sign differs in duration form each other and it is not given to know each "starting point". This problem is, in fact, increased by Movement Epenthesis (ME), i.e. the movement that links two consecutive signs, which is not easily predictable being different each time.

Three different ways of chunking have been investigated in this work:

    a.   Introducing artificial pauses;
    b.   Detecting static configurations;
    c.   Training a neural network to recognize transitions.

The first method is quite trivial. The user is asked to insert a pause before and after they perform the sign, so the chunker detects the sign according to the unchanging measured values.

The second method is somewhat trickier since no artificial pauses are requested. It relies on the fact that in LIS the configuration of the hand (see section III) remains static or replays twice or more times the same gesture during the whole sign. So the new gesture can be identified according to the static or periodically moving hand configuration. This is a simple and efficient way to figure out where to cut the continuous stream of signs, but problems come up when two consecutive signs with the same configuration are performed. In that case the user has to introduce an artificial transition movement (in which fingers are moved disorderly) to split the signs correctly.

The third technique is similar to the one adopted in [7]. Each transition is manually labeled, marking the beginning and the end. A Multilayer Perceptron Neural Network (MLPNN, [7] and [15]) is trained associating the output of the net to a Gaussian centered in the middle of the transition. In this way, the network should learn how to detect a transition, associating an output value in between 0 and 1; outside this range the signal corresponds to a sign (fig. 5).

This method is the most comfortable and natural for the user because does not require them to insert unnatural actions during the "conversation". But unfortunately, the accuracy is lower than the other two methods (about 80% of correct chunking while the former two split correctly almost the totality of the signs) and the learning phase requires a lot of time due to the labeling part.
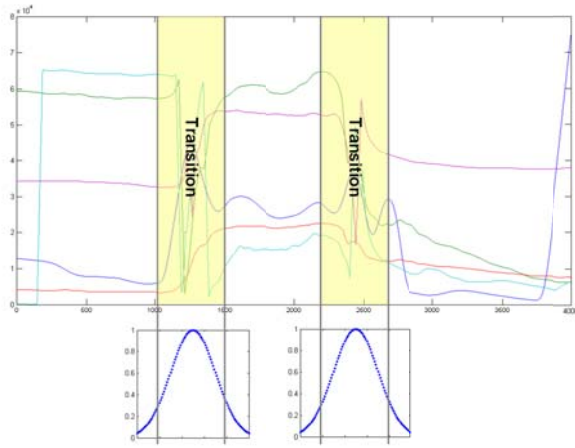
Fig. 5. Labeled transitions and the corresponding neural network's Gaussian output.

The accuracy of the three methods described is reported in table I. A deeper investigation of chunking methods is beyond the purpose of this work; the reader is invited to read papers [8], [9], and [16] for further information.

TABLE I
CHUNKERS ACCURACY

Chunking method	Accuracy on 2000 gestures
a. Artificial pauses	98.8%
b1. Static configurations	92.3%
b2. Static configurations with artificial transitions	97.7%
c. Neural Network	86.4%

Accuracy of chunkers (percentage of correctly chunked sings) evaluated on 200 streams of signs, with an average of 10 signs per stream. Only method b1 and c were tested on the same data set due to the nature of the methods. Method b was tested with and without introducing artificial transition movements.

D. Feature Extraction

Being interested in static postures of the hand we need only to consider one value per each measured joint's angle (besides the space path) to represent the configuration of a sign. Hence, we take into account the median values of the 15 joint angles in order to have enough representative values.

Data from different subjects are calibrated by recording a continuous stream of movements (see section V) to collect the min-max value range of each bend sensor.

The output of the IMUs cannot be directly used as features since it varies according to the duration of the gesture, or the sign related part of the output can be shifted along the time axis.

In order to extract useful features we investigated Wavelet Transform (WT). WT, in contrast to the classic Fourier Transform, works on a multi-scale basis, decomposing a signal into different levels of detail using fixed blocks called wavelets. The wavelet functions used to represent a signal are derived by scaling (dilation) and shifting (translation) a generating function called mother wavelet.

Equation (2) defines the Continuous Wavelet Transform (CWT) of a time signal $x(t)$.

$$W(a, b) = \int_{-\infty}^{+\infty} x(t) \frac{1}{\sqrt{a}} \psi * \left(\frac{t-b}{a}\right) dt \qquad (2)$$

In this equation, ψ is the wavelet function while a and b represent the translation and scale coefficients respectively. The result $W(a, b)$ is the wavelet coefficients that relate the signal to the wavelet at different scales and translations [19].

To reduce the computational workload, a discrete version of the CWT is often used, called Discrete Wavelet Transform (DWT) [20-22]. The DWT is obtained by sampling the time-scale plane using digital filters with different cutoff frequencies at different scales. High-pass filters will produce detailed information, d[n], while low pass filters will give coarse approximations, a[n]. The scale is determined by upsampling and downsampling the signal before filtering it.

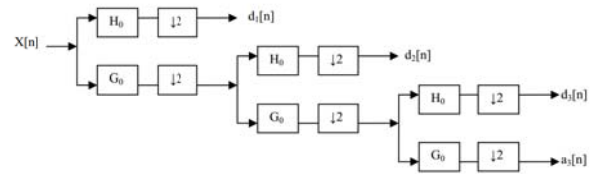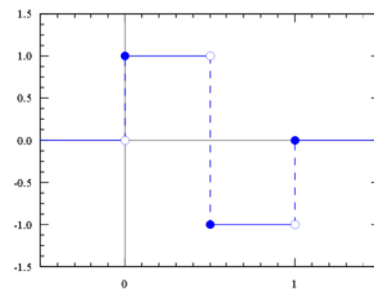The depth of this decomposition tree depends on the

Fig. 6: Wavelet decomposition tree.

length of the signal. The result of the DWT of the time signal is given by the concatenation of all the coefficients a[n] and d[n], from the last level of decomposition.

We performed tests with different types of wavelets,

$$\psi(t) = \begin{cases} 1 & 0 \leq t < 1/2, \\ -1 & 1/2 \leq t < 1, \\ 0 & \text{otherwise}. \end{cases}$$

Fig. 7: Daubechies wavelet of order 2.

and discovered that the Daubechies wavelet of order 2 (fig. 7) gave the maximum efficiency for our application. This is probably due to its smoothing feature that made more suitable to detect changes of the IMU's output values.

After sampling the IMU's signals down to 256 samples, the detail wavelet coefficients at the first, second, third and

fourth levels (129 + 66 + 34 + 18 coefficients) and the approximation wavelet coefficients at the fourth level (18 coefficients) were computed. In order to reduce the dimensionality of the feature vectors, the following statistics over the set of the wavelet coefficients were used to represent each of the six IMU signals, as done in [23]:

(1) Energy of each sub-band.
(2) Maximum coefficient for each sub-band.
(3) Minimum coefficient for each sub-band.
(4) Mean of the coefficients for each sub-band.
(5) Standard deviation of the coefficients for each sub-band.

From our experiments, these 25 features turned out to be sufficient to obtain a good representation of the signal without requiring too much computational power.

Summing everything up, a sign was represented by a feature vector of 315 components (25 features by 6 components for each of the 2 IMUs plus 15 angle values for bending sensors).

E. Classification

The automatic classification of a sign was made by a Support Vector Machine (SVM), in order to find the hyper-plane that maximizes the separation between classes [24].

Let us assume we have two linearly separable classes $y_k \in \{-1, +1\}$. Each training example $\overline{x}^k \in \Re^N$ belongs to a class. A separation hyper-plane between the classes can be written as:

$$y_k(\overline{w}^T \overline{x}^k + b) > 0 \qquad k = 1, ..., m \qquad (3)$$

The aim of an SVM is to maximize the minimum distance between the training examples and the separating hyper-plane, also called margin of separation. We can rewrite (3) rescaling the weights w and the bias b as

$$y_k(\overline{w}^T \overline{x}^k + b) \geq 1 \qquad k = 1, ..., m \qquad (4)$$

Therefore, the margin of separation is $1/\|\overline{w}\|$ and maximizing it is equivalent to minimize the Euclidean norm of the weight vector $\overline{w}$. The optimum hyper-plane will be then found in terms of weights and bias (Fig. 8). All the points $\overline{x}^k$ that satisfy the constraints (4) with the equality sign are called support vectors.

By means of Lagrange Multipliers we are able to consider only these vectors to find the optimal $\overline{w}$ and b. We used a Soft Margin SVM that introduces a tolerance to classification errors. A constant C controls the trade-off between the maximization of the margin and the minimization of the error.

As SVM was originally designed for binary classification, it cannot deal directly with multi-class classification problems, which is usually solved by decomposition of the problem into several two-classes ones. In this context we utilized a *One vs All* strategy [25] where a set of binary classifiers are constructed comparing each time one class to the rest.
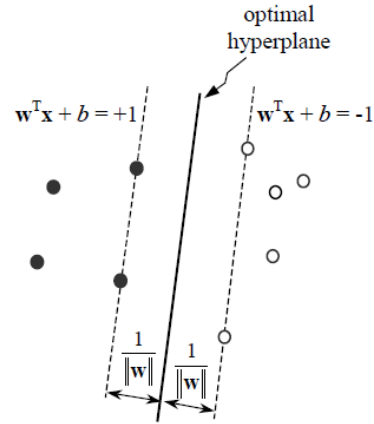

Fig. 8: Optimal separating hyper-plane corresponding to the SVM solution. The support vectors lie on the dashed lines in the caption.

We adopted a linear kernel, and we set the C parameter to a value of 25 through a validation test.

V. EXPERIMENT RESULTS AND DISCUSSION

The experiments were performed with two subjects, the first expert and the second inexpert of the LIS language. The adopted experimental procedure can be summarized as:

- To wear the glove
- To calibrate the system
- To record 100 signs 15 times per sign.

A "session" was the fulfillment of all previous steps. Two sessions for each subject were recorded in two different days in order to test the adaptability of the system to glove repositioning.

Since the piezoresistive sensors are subject to a linear shift of the response function due to temperature change or different repositioning of the glove, a simple calibration was performed each time the user started using the software. The calibration consisted in continuously recording the gesture of opening and closing the hand for ten seconds, in order to find the minimum and maximum values to correctly identify the range of response for each sensor. To avoid noisy outliers the median of several maximum and minimum values was considered instead of one single value.

We trained the SVM with 10 examples of a sign and tested it with the remaining five. The evaluation of the performances of the system was referred to as percentage accuracy according to the following formula:

$$Accuracy = \frac{Correctly\ classified\ examples}{Total\ examples} * 100\% \qquad (5)$$

Table II shows the mean percentage accuracy for each subject, where Subject 1 is the expert LIS user. Tests were done on a growing-size set of gestures consisting of 30, 50

TABLE II
MEAN PERCENTAGE ACCURACY FOR EACH SUBJECT

Subject	Set 30	Set 50	Set 100
I (expert)	96%	95.6%	94.8%
II (non expert)	97.6%	96.4%	93.2%

SVM mean accuracy obtained on a growing size set of signs. Accuracy values are expressed in percentage for each subject. Both training and test set are made from the same session.

and 100 different randomly taken signs, to understand the accuracy trend according to the growth of the dataset.

An average accuracy of 96.8% is obtained on a reduced set of 30 gestures while for the complete set (100 gestures) the accuracy decreases only of 2.8%, obtaining 94% correctly classified signs.

In this first case, we trained and tested the SVM with examples recorded in the same session. It means that the user should train the system every time they wear the glove. This is obviously not convenient and excessively time demanding. In Table III, accuracy achieved training and testing the classifier on different sessions is shown.

TABLE III
MEAN PERCENTAGE ACCURACY ON DIFFERENT SESSIONS

Subject	Session I over Session II	Session II over Session I	Average
I (expert)	88.1%	89.4%	88.75%
II (non expert)	81.2%	80.4%	80.8%

Accuracies obtained training the SVM on a session and testing it on the remaining. Accuracy values are expressed in percentage for each subject.

This is a more practical way since the training phase is performed only once at the very first time, so the user does not have to worry about repositioning the glove.

As we could expect, mean accuracy is lower than the previous test because repositioning the glove affects the overall performance: 84.8% of average accuracy is obtained in this case. However, it has to be noticed that Subject 2 is not an expert LIS user, hence the tendency to forget how to perform a gesture in the second session causes a lower accuracy (80.8%) respect to the expert user (88.75%), but the accuracy remains acceptable.

So far, the system is "ad-personam", in the sense that it has to be trained by the same person who is going to use it.

Table IV reports accuracies when SVM is trained on a

TABLE IV
MEAN PERCENTAGE ACCURACY ON DIFFERENT SUBJECTS

Subject for training	Subject for testing	Mean accuracy
I (expert)	II (non expert)	78%
II (non expert)	I (expert)	73.4%

Accuracies obtained training the SVM on a subject and testing it on the other one. Accuracy values are expressed in percentage.

subject and then tested on the other one. In this case, the system could be trained by an expert user and then used by everyone else, skipping the learning part. Now the system results "weaker" because the two users did not perform signs in the same way, especially if we consider that subject II is a novel LIS user. The average accuracy of 75.7% is reached, meaning that the system is still usable but with a higher error rate, so that it makes sense to integrate an automatic correction system [26].

Although it is not possible to make a comparison with the other A-SLR works reported in literature (because of different acquisition systems and different set of signs), the percentages reached in this study are encouraging. In addition, it has to be considered that our sensory glove is cheaper than state of the art gloves, being equipped with IMUs and not magnetic trackers. The graph in figure 9 represents a rough comparison of production costs for six different types of gesture acquisition systems. As it can be noticed, our system is rather low cost, comparable only to low performance optical systems, which we believe not having the same portability.
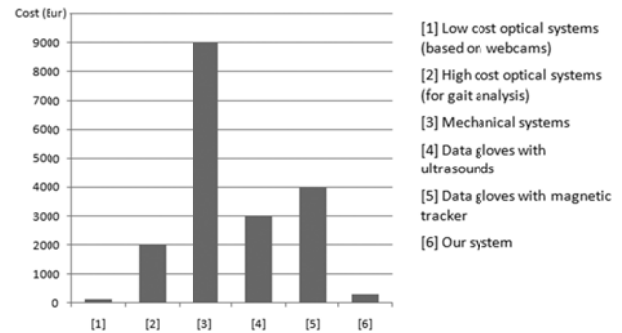

Fig. 9: Approximate costs comparison among different gesture acquisition systems.

VI. CONCLUSION

A completely portable sensory glove was developed and used to recognize 100 gestures belonging to Italian Sign Languag. Based on bend and IMU sensors, the glove is completely portable, unobtrusive and comfortable to don. In the future, our system will become wireless, and the possibility of being interfaced to smartphones will be provided. In this way, it will be easy to carry in any situation, giving the user freedom to move and naturally interact with other people.

An A-SLR system based on wavelet features and SVM is also proposed. Experiments, performed on the same subject, demonstrate results with an average accuracy of 94%. This can be considered a very good result, even though it cannot be directly compared to literature due to the lack of related works about LIS.

In further studies, new ways of chunking and classification will be investigated; the overall system will become wireless and integrated into mobile devices with a speech synthesizer to realize a new aid for everyday communication between deaf and hearing people.

REFERENCES

[1] S.C.W. Ong, S. Ranganath, "Automatic sign language analysis: a survey and the future beyond lexical meaning", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 873-891, Jun 2005.

[2] T. G. Zimmerman et al., "A hand gesture interface device", *SIGCHI Bull.* 17, 1987.

[3] T. Takahashi, F. Kishino, "Hand gesture coding based on experiments using a hand gesture interface device", *SIGCHI Bull.*, 1991.

[4] D. J. Sturman, D. Zeltzer, "A survey of glove-based input", *IEEE Computer Graphics and Applications*, vol. 14, pp. 30-39, Jan 1994.

[5] L. Dipietro et al., "A Survey of Glove-Based Systems and Their Applications", *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews*, vol. 38, pp. 461-482, Jul. 2008.

[6] T. P. Andriacchi, E. J. Alexander, "Studies of human locomotion: past, present and future", *Journal of Biomechanics*, vol. 33, pp. 1217-1224, Mar. 2000.

[7] S. S. Fels, G. E. Hinton, "Glove-Talk: A Neural Network Interface Between a Data-Glove and a Speech Synthesizer", *IEEE Transactions on Neural Networks*, vol. 3, 1992.

[8] G. Fang et al., "Large-Vocabulary Continuous Sign Language Recognition Based on Transition-Movement Models", *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, vol. 37, pp. 1-9, Jan. 2007.

[9] M. Maebatake et al., "Sign Language Recognition Based on Position and Movement Using Multi-Stream HMM", *Second International Symposium on Universal Communication*, ISUC '08, Osaka.

[10] I. Infantino et al., "A System for Sign Language Sentence Recognition Based on Common Sense Context", *The International Conference on Computer as a Tool*, EUROCON 2005.

[11] S. McMillan, "Upper Body Tracking Using the Polhemus Fastrak", *Naval Postgraduate School, Monterey, Technical Report NPSCS-96-002*, Jan. 1996.

[12] Ji-Hwan Kim et al., "3-D Hand Motion Tracking and Gesture Recognition using a data glove", *IEEE International Symposium on Industrial Electronics*, ISIE 2009.

[13] Z. He, "Accelerometer Based Gesture Recognition Using Fusion Features and SVM", *Journal of software*, vol. 6, pp. 1042-1049, Jun. 2011.

[14] S. Mitra, T, Acharya, "Gesture Recognition: A Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, pp. 311-324, May 2007.

[15] D. Xu, "A neural network approach for hand gesture recognition in virtual reality driving training system of spg.", *18th International Conference on Pattern Recognition*, ICPR 2006.

[16] G. Fang, W. Gao, "A SRN/HMM System for Signer-independent Continuous Sign Language Recognition", *Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, Washington DC, USA, 2002.

[17] N. Bertoldi et al., "On the creation and the annotation of a large-scale Italian-LIS parallel corpus", *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, International Conference on Language Resources and Evaluation*, 2010.

[18] O. Romeo, "Dizionario dei segni - La lingua dei segni in 1400 immagini", Zanichelli, 1991.

[19] A. Godfrey et al., "A Continuous Wavelet Transform and Classification Method for Delirium Motoric Subtyping", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 17, no. 3, JUNE 2009.

[20] E. D. Übeyli, "ECG beats classification using multiclass support vector machines with error correcting output codes", *Digital Signal Processing,* vol. 17, pp. 675–684, 2007.

[21] S. Soltani, "On the use of the wavelet decomposition for time series prediction", *Neurocomputing*, vol. 48, pp. 267–277, 2002.

[22] M. Unser, A. Aldroubi, "A review of wavelets in biomedical applications", Proc. IEEE 84, vol. 4, pp. 626–638, 1996.

[23] P. S. Addison, "Wavelet transforms and the ECG: a review", *Physiological Measurement*, vol. 26, pp. 155-199, Aug. 2005.

[24] Vladimir N. Vapnik, "The Nature of Statistical Learning Theory", Springer-Verlag, New York, 1995.

[25] C. W. Hsu, C. J. Lin, "A comparison of Methods for multiclass Support Vector Machine", *IEEE Transactions on Neural Networks*, vol. 13, 2002.

[26] A. S. Dolgopolov, "Automatic spelling correction", *Cybernetics and systems analysis*, vol. 22, pp. 332-339, May 1986.

Pietro Cavallo received the B.S. and M.S. degrees in Computer Engineering from University "Tor Vergata", Rome, Italy in 2008 and 2011, respectively. He collaborates with HITEG group (Health Involved Technical Engineering Group, www.hiteg.uniroma2.it), of Tor Vergata University since 2007. In 2010 he has been an exchange student at University of Bergen, Norway. His research interests include Machine Learning, Human Computer Interaction, Brain Computer Interface, Image Processing and Computer Vision.

Giovanni Saggio received the Dr. Eng. degree in Electronic Engineering from the University "Tor Vergata", Rome, Italy, in 1991, and the Ph.D. degree in Microelectronic and Telecommunication Engineering, in 1996, from the same University. In 1991, he did his thesis in the Nanoelectronics Research Centre, Department of Electronics and Electrical Engineering, University of Glasgow, Scotland, and in the Cavendish Laboratory, Department of Physics, University of Cambridge, England. His initial research activities covered the area of nanodevices, surface acoustic wave devices, noise in electronic devices. He is currently a Researcher at the University "Tor Vergata", Rome, Italy. He teaches courses about Electronics at the Faculty of Engineering and the Faculty of Medicine. His current research interests are related to the fields of biosensors, sensor's characterization, human kinematics' measurements and brain computer interface. He has published tens of papers on international journals and two books about electronics. He is currently: member of Italian Space BioMedical Society; Principal investigator W.P. DCMC Project (from Space Italian Agency); Principal investigator of a Project regarding Aeronautics; Promoter and coordinator of the HITEG group (Health Involved Technical Engineering Group, www.hiteg.uniroma2.it), University "Tor Vergata", Rome, Italy.