

Ultrasound Imaging as a Human-Machine Interface in a Realistic Scenario

Claudio Castellini and David Sierra González

Abstract—Medical ultrasound imaging is a widespread high-resolution (both spatial and temporal) method to gather live images of the interior of the human body. Its potential as a human-machine interface for the disabled — amputees in particular — is being explored in the rehabilitation robotics community.

Following up the recent discovery that first-order spatial features of the ultrasound images of the human forearm are linearly related to the hand configuration, we hereby push the approach to a realistic scenario. We show that an extremely simple calibration procedure can be used to obtain a linear regression system which will effectively predict the forces required by a human subject at the fingertips, using live ultrasound images of the forearm. In particular, the system can be trained on minimum and maximum forces only, thereby dramatically shortening the calibration phase; and it will generalise to intermediate force values. This phenomenon is uniform across 5 intact subjects whom we examined in a controlled experiment. Moreover, it is not necessary to use any force sensor, as learning-by-imitation, namely using a visual stimulus, yields similar results.

This result is particularly useful in the case of amputees, who normally cannot perform graded-force tasks as proprioception may be lost since decades. Applications of this system include, among others: advanced prosthetics, phantom pain therapy and smart teleoperation.

I. INTRODUCTION

Medical ultrasonography (US) is a non-invasive technique to visualise structures inside the human body [1], exploiting the principle of wave reflection. Piezoelectric transducers are used to generate a focused wave of ultrasound which penetrates the body part of interest; partial reflection of the wave at the interfaces between tissues with different acoustic impedance is converted to a grey-scale 2D image (in the so-called B-mode). High values of grey denote tissue interfaces. Modern US machines are portable or even hand-held, can be as cheap as 5000EUR and can achieve sub-millimeter spatial resolution and 40Hz temporal resolution, penetrating several centimeters below the subject's skin. US imaging has no known side effects [2] and is routinely used in all hospitals.

Recently, extensive work by Zheng [3], [4], [5] and ourselves [6], [7] has revealed that US imaging can also be used as a human-machine interface (HMI). In [6] in particular, we have for the first time shown that a linear relationship exists between the angles at the metacarpophalangeal joints of the human hand and spatial first-order features extracted from

The authors would like to acknowledge Mr. Tilo Wüsthoff of the DLR for designing the transducer cradle.

The authors are with the Robotics and Mechatronics Center, DLR — German Aerospace Center, D-82234 Oberpfaffenhofen, Germany {claudio.castellini,david.sierragonzalez}@dlr.de

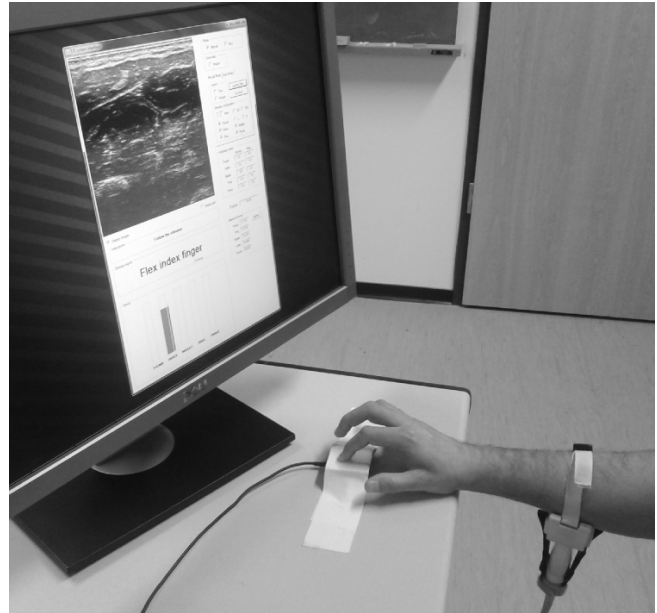


Fig. 1. A bird's eye view of the setup. The subject sits in front of a screen on which the stimulus is shown as vertical coloured bars; meanwhile force data and US images are recorded.

US images of the forearm. Since the metacarpophalangeal joints are those at the basis of the fingers (linking each finger to the palm), it is possible to reconstruct the hand configuration using the US images of the forearm. As the information is strictly positional, the system works fine irrespective of the velocity of the subject's movement, the only limitation being the imaging rate of the ultrasound machine. In a nutshell, US images can be used to reconstruct the kinematic configuration of a human hand. The only piece of machinery which must be in contact with the subject is the US transducer placed on the forearm, thus leaving the hand completely free (no obstructions from sensors) to perform any movement. Interesting potential applications range from teleoperation and hand prosthetics (as a very ergonomic and precise HMI) to visualising the imaginary limb of an amputee, or of a nerve-injury patient.

In this paper we propose a further, more detailed and more realistic analysis of the possibilities given by US imaging as an HMI. In particular, we show that (a) a linear relationship also exists between spatial first-order US image features and *forces at the fingertips*; and that (b) to build this relationship it suffices to gather data from a human subject *only when resting and exerting maximum force*; the model will then be able to correctly predict the intermediate force values, too.

Predicting forces rather than positions, as it was the case in [6], [7], is probably even more useful, since a prosthetic hand (and a hand, indeed!) will usually be employed in grasping objects, therefore benefiting from force control more than position control. Moreover, that a simple "on-off" training generalises to graded force values has at least two desirable implications: firstly, a much shorter training time is required, with respect to what is shown in our previous work; secondly, the system is potentially applicable to amputees, who in the general case retain no proprioception and are therefore unable to perform graded-task forces with the imaginary limb. This latter possibility goes in the direction of phantom pain treatment: it is well-known since the 90s [8], [9] that restoring the visual feedback loop from the missing limb has beneficial effects.

We substantiate these claims by showing the results of a graded-force task exercise performed by 5 intact subjects, who train the system by only resting and applying maximum force. The linear predictor so obtained achieves a normalised root-mean-squared-error of 10% on graded tasks, a result which is in line with previous studies. The results are uniform across subjects and fingers. As an aside result, we show that the training/testing works as well when only a visual stimulus is used as the ground truth, thereby eliminating the need of any force sensors whatsoever. This is all the more useful in the case of amputees, when position/force ground truth is not available in principle. An amputee could train the system simply by following a simple on-off pattern appearing on a PC screen.

The paper is organised as follows: Section II describes the experiment; Section III describes the data analysis and experimental results, and lastly Section IV draws some conclusions.

II. EXPERIMENT DESCRIPTION

A psychophysical experiment was set up in order to prove that (a) force at the fingertips can be linearly predicted using ultrasound images, and (b) an on-off training schema (that is, only resting and applying maximum force) is enough to predict graded forces.

A. Ultrasound imaging

Ultrasound images were gathered using a General Electric *Logiq-e* portable ultrasound machine (www.gehealthcare.com) equipped with a 12L-RS linear transducer. The machine is set to B-mode, resulting in a gray-valued image representing a section of what lies directly under the probe. Similarly to [7], we chose the following settings: ultrasound frequency of 12MHz, edge enhancement on, focus point at a depth of about 1.3cm, minimum depth of field. This results in a frame rate of 38Hz.

Movement of the probe with respect to the subject's skin, which would have severely hampered the system (see [6], [7] again) was avoided using a custom-built plastic cradle obtained via rapid prototyping. The cradle hosts the transducer's head on one side (velcro straps attach the transducer to the cradle), while being lightly but firmly tied to the

forearm on the other side by means of a biocompatible elastic band and a side-release buckle. Figure 2(a,b,c) show the transducer, the cradle and the transducer+cradle fixed on a subject's forearm.

After extensive initial visual checks, we fixed the transducer on the ventral side of the forearm, at a distance of about 10cm from the elbow. The typical output image (consider Figure 3) contains the ulna and the main flexor muscles and tendons. The images are captured from the ultrasound machine's VGA video output using a commercial PCIe video capture card, running at 60 frames per second. As the frames are captured asynchronously with respect to the ultrasound machine, not all of them are whole ultrasound images. In order to avoid considering torn or repeated frames, we enforced the same kind of filtering of [7], obtaining a valid frame rate of slightly less than 30 frames per second.

B. Fingertip forces

An ATI Mini45 SI-290-10 force sensor (www.ati-ia.com) with guaranteed linear output and a resolution of $\frac{1}{8}$ N was used to gather forces at the fingertips. The sensor was taped onto the setup table at a convenient distance from the subject's hand, so that a minimal movement would be involved in pressing it with each finger. The sensor was connected to a DAQ card, and its values read and streamed over on a UDP stream on the local network. Figure 2(d) shows the force sensor.

C. Experiment subjects and protocol

Five healthy, intact human subjects (age 29.8 ± 6.8 , max 40, min 23, all right-handed) joined the experiment. All of them received a thorough description of the experiment, both in oral and written form, and signed an informed consent form prior to starting. Each subject was then asked to press the sensor once with each finger, applying the largest possible force without feeling discomfort or pain. This way we gathered an indication, F_{max} , of the maximum forces applicable by each subject at each finger.

At the beginning of the experiment, each subject sat comfortably on an adjustable office chair, maintaining an upright body posture with both feet on the floor and the elbow bent at 90° . The subjects were then asked to simply lean their dominant hand on the table next to the sensor and, during the experiment, do as instructed by the stimulus displayed on a computer screen. Figure 1 shows a bird's eye view of the setup.

The experiment consisted of two identical sessions, and each session was likewise divided in two parts: an *on-off* phase (OO) and a *graded* phase (GR). In the OO phase, the stimulus induced the subject to either rest or apply maximum force with each finger; in the GR phase, the subject was induced to exert forces following a squared sinusoidal pattern, i.e., to apply a full range of forces from none to maximum. The complete structure of the stimulus for one of the sessions is displayed in Figure 4(a). The different phases will be hereafter denoted as OO1 and GR1 (for session 1) and OO2 and GR2 (for session 2).

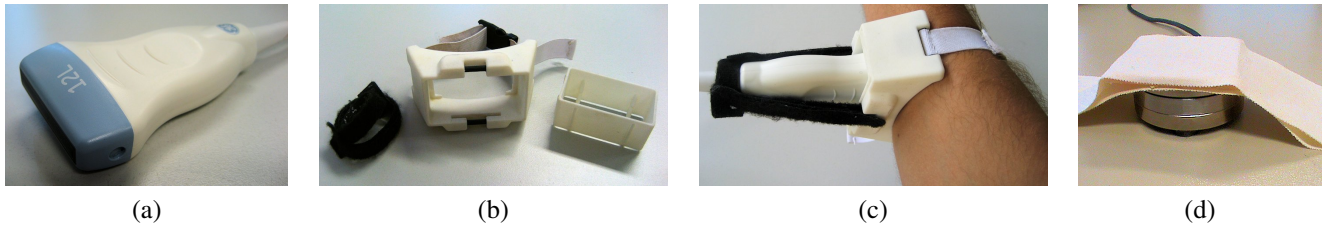


Fig. 2. Parts of the setup: (a) the linear ultrasound transducer GE 12L-RS; (b) the custom-made transducer cradle; (c) the transducer fixed onto a subject's forearm, using the cradle; (d) the ATI Mini45 force sensor, taped to the table. The subjects press on its top.

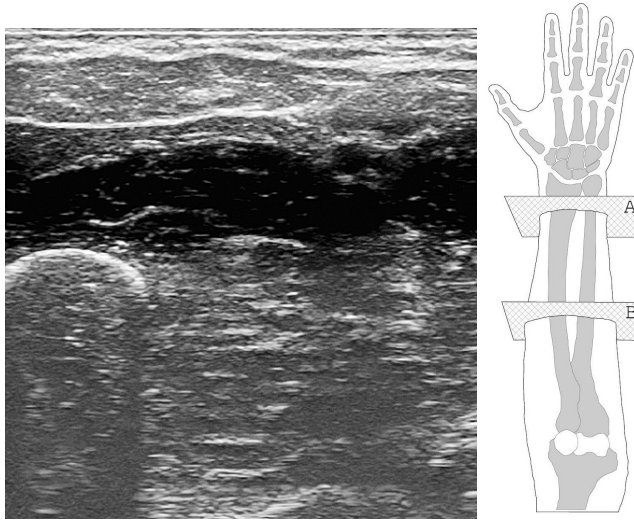


Fig. 3. (left) A typical ultrasound image obtained during our experiment. The ulna is clearly visible in the bottom-left corner, while the flexor muscles and tendons are seen in the upper part. (right) A graphical representation of the human forearm and hand (right forearm; dorsal side up). The transducer is placed onto the *ventral* side; plane "B" corresponds to the section of which the transducer would return an image.

In the on-off phases (OO1 and OO2), the stimulus consisted of a series of large banners and vertical coloured bars appearing on the screen, representing the force pattern to be applied; e.g., "rest", "press with the index finger" or "press with the thumb". (Notice that pressing with the thumb in this configuration is tantamount to rotating it, for example when hitting a piano key with the thumb.) The subject was instructed to press with the required finger on the sensor applying "a reasonably large amount of force". This intentionally fuzzy indication reflects what can be asked of an amputee.

In the graded phases (GR1 and GR2), two coloured bars were displayed on the screen: one bar representing the required force, one showing the force actually applied at the sensor's surface. The stimulus for the required force was chosen as $0.8F_{max} \sin^2(t)$. Figure 4(b) shows the force measurements of the sensor for a typical subject during one of the sessions.

During the on-off phases the subject was told to rest or apply maximum force with each finger in turn (little, ring, middle, index and thumb), and the whole cycle was repeated 5 times. Each flexion lasted 4.5s, and 4.5s of rest were allowed in-between flexions. Additionally, the transition of

the coloured bars from rest to maximum force and vice versa lasted 1s each. This results in a duration of $5 \times 5 \times (4.5s + 1s + 4.5s + 1s) = 275s$ for each on-off phase. In the graded phases the same sequence was administered; in this case each pattern (from rest, increasing the force to maximum then decreasing again to rest according to the \sin^2 pattern) lasted again 4.5s, and 1.5s of rest was allowed in-between flexions. This results in $5 \times 5 \times (4.5s + 1.5s) = 150s$ for each graded phase.

All in all the experiment lasted $275s + 150s + 275s + 150s = 850s = 14'10s$. No subjects reported discomfort of fatigue during or after the experiment.

D. Data preprocessing

All data was collected on a Windows PC. Visual inspection revealed that there was no measurable delay during the UDP transmission, which enabled us to use the valid frame rate, 30Hz, as the global sampling frequency. Notice anyway that the bandwidth of the signals we are interested in (i.e., frames and force data) is directly dependent on the stimulus, that is way less than one Hertz.

From each ultrasound image the same kind of features used in [7] were extracted; namely, 181 uniformly distributed circular regions of interest (ROIs) of radius 20 pixels were selected on the image. Let the i th ROI be centered around (x_i, y_i) ; then from each ROI three real numbers $(\alpha_i, \beta_i, \gamma_i)$ were computed, such that the grey values of each pixel $(x, y) \in \text{ROI}_i$ would be approximated by $\alpha_i(x-x_i) + \beta_i(y-y_i) + \gamma_i$. Intuitively, α_i denotes the mean image greyscale gradient along the x direction (rows of the image), β_i is the same value along the y (columns) direction, and γ_i is an offset. The three features represent a first-order spatial approximation of the grey values of the ROI, accounting for the morphological structure of that region. In order to extract these features (and for all other image-related computations and evaluations) we used the HALCON v10.0 library by MVTec (see www.mvtec.com/halcon). Since three numbers were extracted from each ROI, the dimension of one US sample is 543.

All signals (force and image features) were lowpass filtered with a Butterworth first-order low-pass filter, cutoff frequency of 1Hz. From the data in the on-off phases only the last two thirds of the on and off periods were taken into account in order to avoid considering the transitions from rest to maximum force and vice-versa.

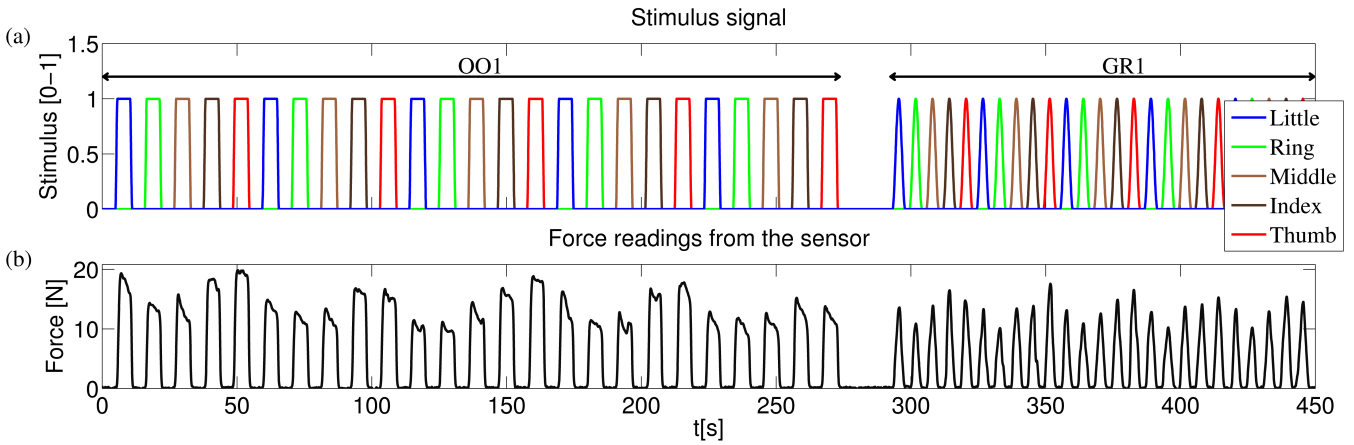


Fig. 4. (a) Structure of the stimulus shown to the subjects, first session. In the on-off phase (OO1), only rest and maximum force are induced for each finger, each repetition consisting of 4.5s of force application, followed by 4.5s of rest. Five repetitions per finger are induced. In the graded phase (GR1) the subjects must exert force following a squared sinusoidal pattern. (b) Forces as measured by the force sensor during the experiment for a typical subject.

III. DATA ANALYSIS AND EXPERIMENTAL RESULTS

In this work we are particularly interested in a *realistic* scenario in which to conduct an experiment and the corresponding analysis. Thus, we will describe the analysis separately according to which signal is considered the ground truth: either the force, as recorded by the force sensor, or the stimulus. In the first case we consider the force applied by each finger during the flexion, and zero force otherwise; in the second, the stimulus itself is used, with the hope that the subject has followed it with a certain degree of precision. Notice that this second scenario reflects the typical situation with an amputee, in which *no ground truth is available in principle* and one must resort to either imitation learning or bilateral action (see, e.g., [10], [11]) — using a visual stimulus and instructing the subjects to imitate it is tantamount to imitation learning.

We hereby try and extend to forces the phenomenon first described in [6], [7], according to which a *linear relationship* exists between the ultrasound image features described in the previous Section and the angles at the metacarpal hand joints. Accordingly, we try and associate the feature vector $\mathbf{v} \in \mathbb{R}^{543}$ extracted from each frame to a finger force $f_i \in \mathbb{R}$ or, alternatively, to a stimulus value $s_i \in \mathbb{R}$, where $i = 1, \dots, 5$ denotes the fingers, in turn the little, ring, middle, index and thumb:

$$f_i = \mathbf{w}_f^T \mathbf{v}$$

$$s_i = \mathbf{w}_s^T \mathbf{v}$$

In order to find the optimal \mathbf{w}_f and \mathbf{w}_s we use the standard technique called *ridge regression*, essentially a regularised form of least-squares regression: given a matrix and vector containing the sample and target pairs (X, \mathbf{y}) , the weight vector \mathbf{w} for which $\mathbf{y} = \mathbf{w}^T \mathbf{x}$ holds is given by

$$\mathbf{w} = (X^T X + \lambda I_d)^{-1} X^T \mathbf{y}$$

where d is the dimension of the input space, I_d is the identical matrix of order d and $\lambda > 0$ is the regularisation

coefficient, which we consistently set at the standard value of 1. Notice that the evaluation of \mathbf{w} involves inverting a matrix of dimension $d \times d$, regardless of the number of samples in the dataset considered.

The feasibility of the linear approximation is checked by considering the square-root mean-square error normalised over the range of the target values (nRMSE), between the measured force values or stimuli and the approximation.

A. Prediction of forces using US images

In order to check whether US image features can be used to predict finger forces, we first used data captured during each phase separately (OO1, GR1, OO2 and GR2), employing cross-validation on each dataset. For each phase we used a randomly chosen 10% of the data to estimate \mathbf{w}_f and \mathbf{w}_s ; we then evaluated the forces and stimulus values on the remaining 90% of the dataset. This procedure was repeated for 50 times (each time with a different training set, obviously) and then the mean and standard deviation of the nRMSE were considered. Figure 5 shows the obtained nRMSE values for a typical subject, according to the finger and the chosen ground truth (force or stimulus).

This analysis was repeated for all subjects. Figure 6 shows the obtained nRMSE values for all subjects altogether.

As is apparent from the Figures (consider especially Figure 6), the linear regression is able to approximate all required values to a remarkable precision. All on-off values are predicted with a nRMSE of 1% of the force ranges or less. Graded phases (GR1 and GR2) exhibit a higher error, slightly higher than 1.5% in case the force is used as ground truth, and slightly higher than 2% in case the stimulus is used. These results are consistent across subjects and fingers. We believe this is reasonable, since in the graded case many more different values must be predicted; moreover, in case the stimulus is used as ground truth, there is an inevitable discrepancy between the stimulus and the actual action performed by the subject. This increases the uncertainty.

Notice that these error levels are obtained by training

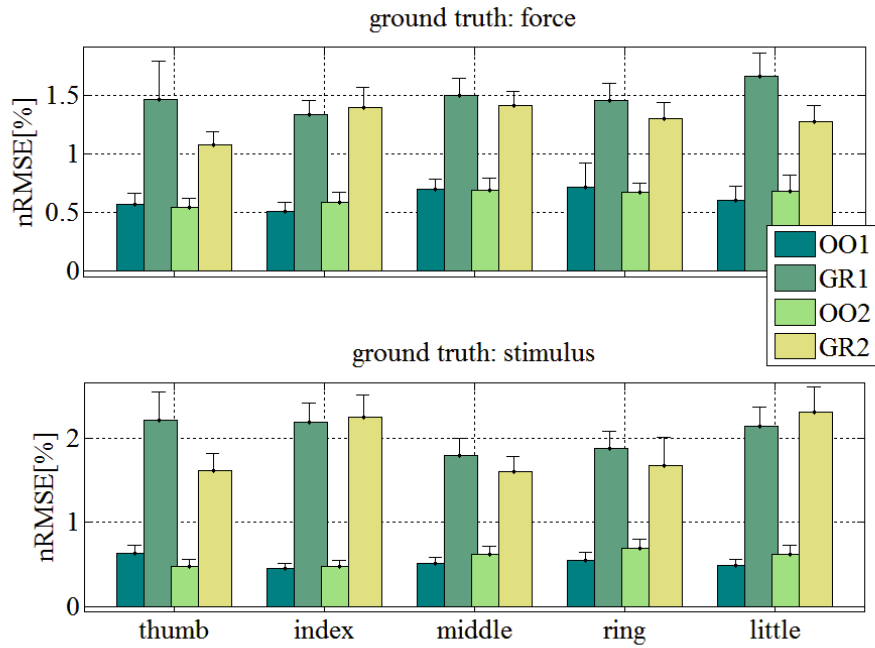


Fig. 5. Results for a typical subject, for each phase (OO1, GR1, OO2 and GR2) and finger. Each bar and stem represent the mean and standard deviation of the nRMSE obtained on the 50 cross-validation folds considered. (top panel) Errors obtained with the force as ground truth; (bottom panel) errors obtained with the stimulus values as ground truth.

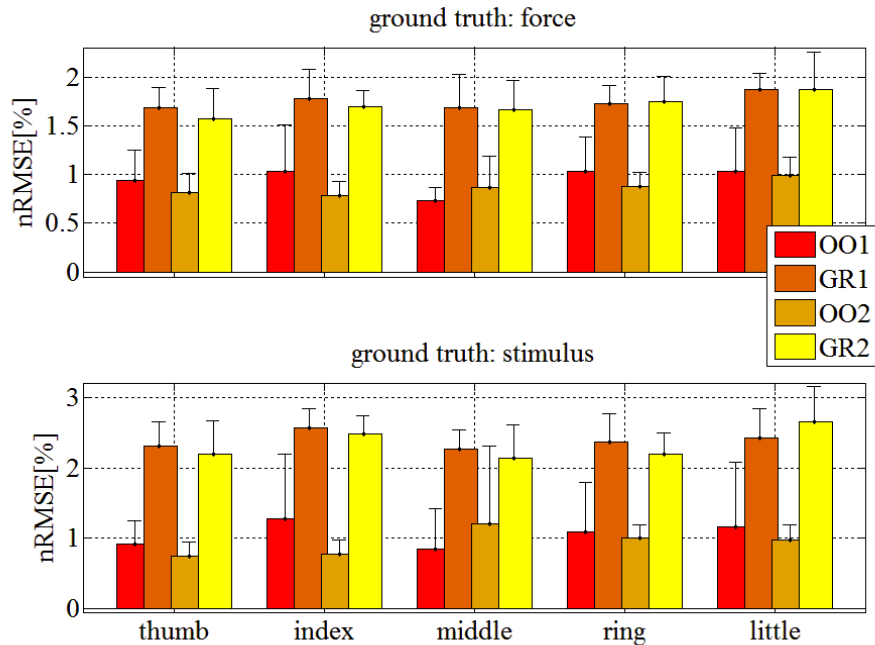


Fig. 6. Results for all subjects, for each phase (OO1, GR1, OO2 and GR2) and finger. Each bar and stem represent the mean and standard deviation of the mean nRMSE values obtained for each subject.

on one tenth of the available data, and are comparable to those presented in [7]. From this we conclude that *a linear relationship exists between finger forces and ultrasound images.*

B. Prediction of graded forces using the on-off data

An even more interesting question is whether the linear relationship discovered between finger forces and ultrasound images is able to generalise to graded forces. That is, given

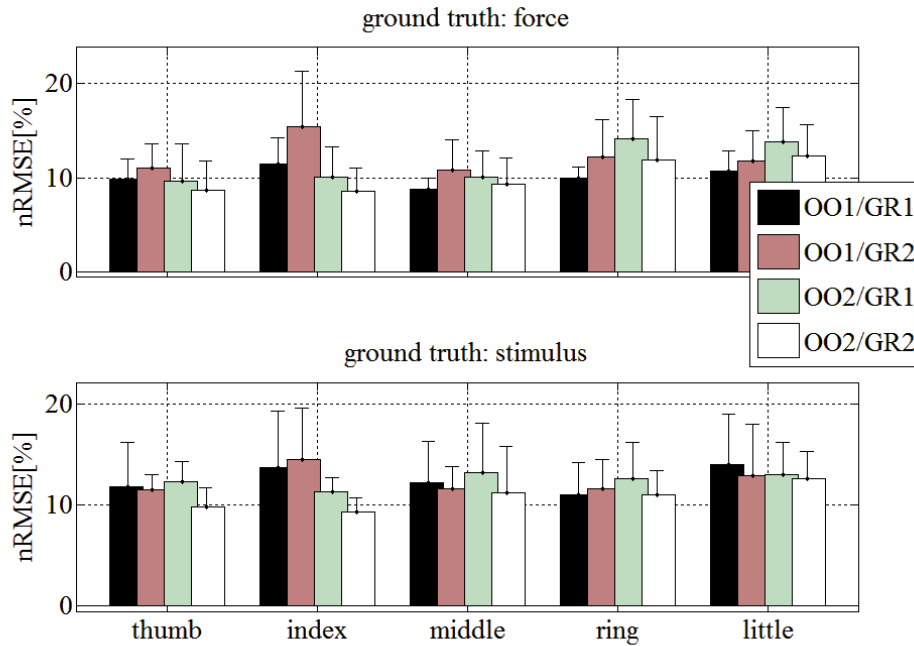


Fig. 7. Results for all subjects, when training on an on-off phase and testing on a graded phase. The legend denotes the training/testing phase, e.g., OO1/GR2 means that the machine was trained on data gathered during the first on-off phase and tested on data gathered during the second graded phase. (top panel) Errors obtained with the force as ground truth; (bottom panel) errors obtained with the stimulus values as ground truth. Each bar and stem represent the mean and standard deviation of the mean nRMSE values obtained for each subject.

only data obtained while resting and applying maximum force, will the system be able to correctly predict intermediate force values? In order to find an answer we estimate the w_f and w_s on the on-off phases (OO1 and OO2) and then test the performance of the linear approximant on the graded phases (GR1 and GR2). The results for all subjects altogether are visible in Figure 7.

The overall nRMSE is clearly much larger than in the previous case, this time around 10% of the target range. Notice, however, that this error is remarkably consistent over subjects, fingers and chosen training/testing datasets (i.e., there is no statistically significant difference when OO1 or OO2 is used to estimate the linear regressors, as well as there is no difference when testing on GR1 or on GR2). This error level is, again, comparable with that obtained in [7]. Also, by comparing the upper and lower panels of Figure 7, it is apparent that there is no big difference when using the force as ground truth, or the stimulus itself.

Figure 8 shows some typical force and stimulus values, predicted on a graded phase (GR2) using a linear predictor trained on an on-off phase (OO2). A non-uniform delay of 300-500ms is visible in the lower panel, where the stimulus is used as ground truth (overall nRMSE 8.27%); this is to be expected because of the inevitable delay introduced by each subject while following the stimulus.

IV. CONCLUSIONS

This work advances the state of the art towards the usage of ultrasound imaging as an HMI. In particular, we hereby

pose the following two questions: (a) can we predict finger forces from the ultrasound images of the forearm? and (b) can we limit the data collection to minimum and maximum finger flexion forces, and still obtain a good approximation of graded forces? In order to answer, we have collected data from 5 intact subjects while applying "on-off" and "graded" force patterns to a high-accuracy force sensor. US images, force and stimulus values were synchronously recorded during the experiment.

The answer to both questions is positive. In detail: the experimental results presented in Subsection III-A show that the US features already used in [7], namely first-order spatial local approximants of the grey levels, are strongly linearly related to the forces at the fingertips, leading to a prediction error always smaller than 1% in the on-off case and between 1% and 1.5% in the graded case. In case the *stimulus* is used as ground truth, instead of the force values recorded by the sensor, the error increases only slightly, namely up to about 2%.

As far as the second question is concerned, an error rate of about 10% nRMSE is obtained uniformly across subjects and fingers, whenever on-off data is used to predict graded forces. This happens regardless whether the force or the stimulus is used as ground truth. This error level is consistent with the existing literature. We hypothesise that this is due to the inherent properties of a linear relationship: knowing the values of a target in two points of interest is enough to interpolate all remaining values.

These results pave the way to the use of US imaging as

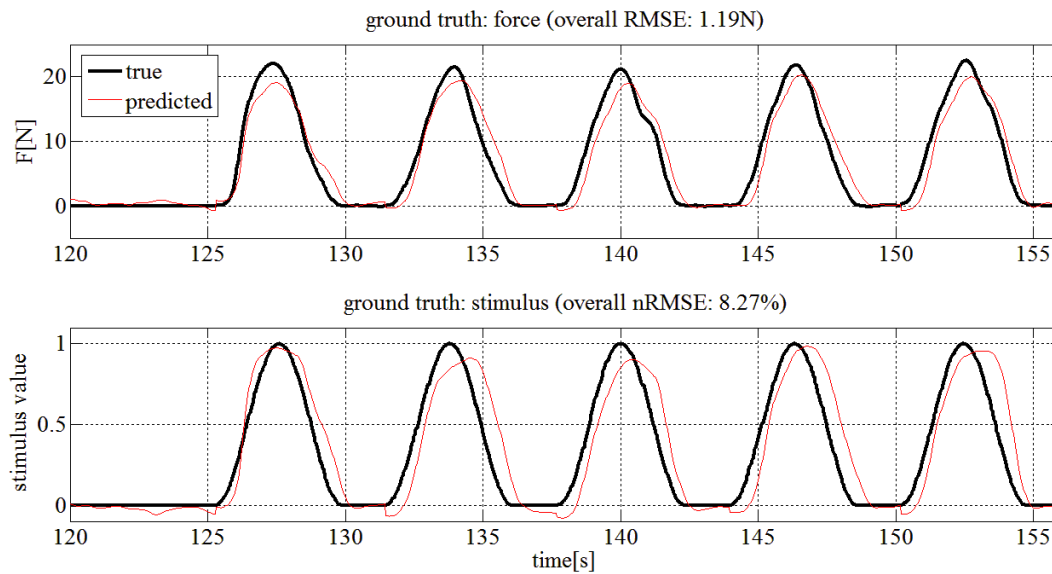


Fig. 8. Typical recorded and predicted signals. Training on OO2, testing on GR2.

a fully fledged HMI for the disabled. In particular we claim that this system will be able to visualise the imaginary limb of an amputee (whenever enough residual muscular activity is present in the stump), as well as to constitute a novel, totally non-invasive and ergonomic HMI for intact subjects in, e.g., teleoperation of virtual reality. Remarkably, the prediction is done using a linear predictor (ridge regression), which can be easily made incremental and put online. The fact that the system can be trained on on-off patterns only and using just a visual stimulus, i.e., that no force sensors are required, makes the usability in the case of amputees very high. We plan to check if this system can in particular be used as an enhanced form of mirror therapy [8], [9] to treat phantom-limb pain.

REFERENCES

- [1] Richard S. C. Cobbold. *Foundations of Biomedical Ultrasound*. Biomedical Engineering. Oxford University Press, 2007.
- [2] *Training in diagnostic ultrasound: essentials, principles and standards: report of a WHO study group*. WHO Technical Report Series, nr. 875. World Health Organisation, 1998.
- [3] Y.P. Zheng, M.M.F. Chan, J. Shi, X. Chen, and Q.H. Huang. Sonomyography: Monitoring morphological changes of forearm muscles in actions with the feasibility for the control of powered prosthesis. *Medical Engineering & Physics*, 28:405–415, 2006.
- [4] Xin Chen, Yong-Ping Zheng, Jing Yi Guo, and Jun Shi. Sonomyography (SMG) control for powered prosthetic hand: a study with normal subjects. *Ultrasound in Med. & Biol.*, 36(7):1076–1088, 2010.
- [5] G. Jing-Yi, Z. Yong-Ping, L. P. J. Kenney, A. Bowen, D. Howard, and J. J. Canderle. A comparative evaluation of sonomyography, electromyography, force and wrist angle in a discrete tracking task. *Ultrasound in Med. & Biol.*, 37(6):884–891, 2011.
- [6] Claudio Castellini and Georg Passig. Ultrasound image features of the wrist are linearly related to finger positions. In *Proceedings of IROS - International Conference on Intelligent Robots and Systems*, pages 2108–2114, 2011.
- [7] Claudio Castellini, Georg Passig, and Emanuel Zarka. Using ultrasound images of the forearm to predict finger positions. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20(6):788–797, 2012.
- [8] V. S. Ramachandran, D. Rogers-Ramachandran, and S. Cobb. Touching the phantom limb. *Nature*, 377(6549):489–490, 1995.
- [9] Brenda L. Chan, Richard Witt, Alexandra P. Charrow, Amanda Magee, Robin Howard, Paul F. Pasquina, Kenneth M. Heilman, and Jack W. Tsao. Mirror therapy for phantom limb pain. *New England Journal of Medicine*, 357(21):2206–2207, 2007.
- [10] Claudio Castellini, Emanuele Gruppioni, Angelo Davalli, and Giulio Sandini. Fine detection of grasp force and posture by amputees via surface electromyography. *Journal of Physiology (Paris)*, 103(3-5):255–262, 2009.
- [11] J.L.G. Nielsen, S. Holmgaard, Ning Jiang, K.B. Englehart, D. Farina, and P.A. Parker. Simultaneous and proportional force estimation for multifunction myoelectric prostheses using mirrored bilateral training. *Biomedical Engineering, IEEE Transactions on*, 58(3):681–688, March.